

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TECNICA SUPERIOR DE INGENIERÍA  
AGRONÓMICA ALIMENTARIA Y DE BIOSISTEMAS



ESPECTROSCOPIA NIR EN ACEITUNA DE ALMAZARA:  
EXPLORACIÓN DE MÉTODOS QUIMIOMÉTRICOS PARA  
ESTIMACIÓN DE CALIDAD

MASTERANDO  
MARY EUGENIA TURPO RODRÍGUEZ

TURORA

BELÉN DIEZMA IGLESIAS (DOCTORA INGENIERA  
AGRÓNOMA)

MADRID 2016

## TABLA DE CONTENIDO

1	RESUMEN .....	11
2	INTRODUCCIÓN .....	14
3	REVISIÓN BIBLIOGRÁFICA.....	15
3.1	PRODUCCION OLEICA.....	15
3.1.1	Descripción de la planta de olivo .....	16
3.1.2	Problemas sanitarios en los cultivos de olivo.....	16
3.2	Parámetros fisicoquímicos de calidad.....	19
3.2.1	Composición del aceite de oliva.....	20
3.2.2	Problemática con el control de calidad de fruto para la producción oleícola.....	20
3.3	ESPECTROMETRIA NIR.....	21
3.3.1	Equipos espectrómetros.....	25
3.3.2	Aplicaciones .....	26
3.4	MÉTODOS QUIMIOMÉTRICOS.....	27
3.4.1	Quimiometría .....	27
3.4.2	Pretratamiento espectral .....	27
3.4.3	Métodos de Reconocimiento de Pautas.....	28
4	PLANTEAMIENTO DEL PROBLEMA .....	32
5	ANTECEDENTES Y OBJETIVOS.....	33
5.1	Antecedentes .....	33
5.2	Objetivos .....	34
6	METODOLOGÍA EXPERIMENTAL.....	35
6.1	Colectivo Muestral .....	35
6.2	Instrumentación .....	37
6.3	Software .....	38
6.4	Métodos y análisis de referencia .....	38
6.4.1	Toma espectral .....	39
6.4.2	Determinación del contenido en humedad .....	39
6.4.3	Análisis de datos.....	40
7	RESULTADOS Y DISCUSIÓN.....	42
7.1	Variedad Picual.....	42
7.1.1	Evaluación por categorías Primera (P), Segunda (S) y Tercera (T). .....	44
7.1.2	Evaluación por categorías: Buenas (B) y Dañadas (D). .....	48
7.1.3	Rendimiento graso.....	51
7.1.4	Contenido de humedad.....	52

7.2	Variedad Cornicabra.....	53
7.2.1	Evaluación por categorías Primera (P), Segunda (S) y tercera (T). ....	53
7.2.2	Evaluación entre categorías de frutos buenos (B) y dañados (D).....	57
7.2.3	Rendimiento graso.....	60
7.2.4	Contenido de humedad.....	60
8	CONCLUSIONES .....	63
9	REFERENCIAS.....	64
10	ANEXOS .....	67

## INDICE DE FIGURAS

Figura 1 Evolución de la superficie de aceituna de almazara en España (ha), Fuente MAGRAMA (2013) .....	15
Figura 2 Evolución de la producción de aceite de oliva en España (t), Fuente MAGRAMA (2013) .....	15
Figura 3 Izquierda: Daño de larva en fruto de oliva (Izquierda - P. del Estal Padillo, Unidad de Protección de Cultivos, E.T.S.I. Agrónomos, Universidad Politécnica de Madrid y Sección de Agricultura y Sanidad Vegetal, D.G. Agricultura y Desarrollo Rural, Comunidad de Madrid – derecha). .....	17
Figura 4 Daño de polilla en fruto del olivo .....	18
Figura 5 Incidencia de cochinilla violeta en frutos de aceituna (imagen de AgroEs.es) .....	18
Figura 6 Rama de olivo infestada con tuberculosis.....	19
Figura 7 Arreglo del Espectro electromagnético; Fuente: Scheme of Electromagnetic spectrum (University of Prince Edward Island. ....	22
Figura 8 Configuración para la adquisición de ( a ) la reflectancia , ( b ) la transmitancia , y los espectros interactancia ( c ) , con ( i ) la fuente de luz , ( ii ) la fruta , ( iii ) monocromador / detector, ( iv ) la barrera de luz , y ( v ) de apoyo. En el modo interactancia, la luz debido a la reflexión está físicamente impedido de entrar en el monocromador por medio de una barrera de luz ( iv ) . $I_0$ es la intensidad de la luz original, $I$ es la intensidad de la luz medida por el monocromador / detector (Nicolai et al., 2007). ....	23
Figura 9 Componentes básicos de un instrumento NIRS; (Salguero-Chaparro et al., 2013) .....	26
Figura 10 Mezcla inicial de aceitunas por categorizar .....	35
Figura 11 Categorías de aceitunas; verde y pintonas (P), moradas (S) y negras (T) de izquierda a derecha.....	36
Figura 12 Colectivo muestral de aceitunas .....	36
Figura 13 Equipamiento electrónico del espectrómetro .....	38
Figura 14 (Izquierda) Toma espectral y vista del espectrofotómetro, (Derecha) acercamiento de la toma espectral.....	39
Figura 15 Muestras secadas en horno .....	40



Figura 16 Loading correspondiente a la aplicación de ACP sobre los espectros no tratados de aceitunas frescas cosechadas hasta el 21 de diciembre. ....	42
Figura 17 Proyección de las aceitunas frescas (cuadrados rojos) y deshidratadas (resto de puntos), categorizados según la agrupación del análisis de clústeres.....	43
Figura 18 Espectros de aceitunas deshidratadas de la variedad Picual sin pretratamiento, Derecha: Ídem, normalizados. ....	44
<i>Figura 19 Loadings para espectros sin pretratamiento con valores elevados en torno a 1.200 y 1.400 nm – Variedad Picual.....</i>	<i>44</i>
Figura 20 Loading para PCA con espectros normalizados – Variedad Picual.....	45
Figura 21 Scores de los PC4 y PC5 de los espectros sin pretratamiento – Variedad Picual.....	45
Figura 22 Gráfico de Cajas;( Izquierda) para el score 5 sobre las muestras crudas; (Derecha) Ídem sobre los datos normalizados – Variedad Picual.....	46
Figura 23 (Izquierda) Comparación múltiple de los datos estadísticos del ANOVA sobre el PC5 de los espectros crudos para determinar si existen diferencias significativas entre las categorías P, S y T; (Derecha) Ídem sobre los espectros normalizados – Variedad Picual.....	46
Figura 24 Izquierda: Gráfico de cajas para el análisis PLSDA con espectros crudos; Derecha: Gráfico de cajas para el análisis PLSDA sobre los espectros normalizados – Variedad Picual....	47
Figura 25 Izquierda: Análisis de Comparación múltiple para los scores obtenidos mediante PLSDA sobre los espectros crudos para las categorías P, S y T; Derecha. Ídem para espectros normalizados. – Variedad Picual .....	48
Figura 26 Loading correspondiente a la aplicación de ACP sobre los espectros no tratados de aceitunas frescas – Variedad Picual .....	48
Figura 27 Gráfico de cajas del ANOVA Realizado sobre los scores del PC5 para determinar la significancia entre las categorías B y D – Variedad Picual.....	49
Figura 28 Vector de proyección para el análisis PLSDA – Variedad Picual.....	49
Figura 29 Gráfico de cajas para el ANOVA realizado sobre los scores del PLSDA, para determinar la significancia entre B y D – Variedad Picual.....	50
Figura 30 Scores de los PC5 y PC7 de los espectros sin pretratamiento de las categorías B y D – Variedad Picual.....	51
Figura 31 (Izquierda) ANOVA del PC5 para determinar la significancia entre los días de recolección de las aceitunas de almazara;(Derecha) Representación de los días de recolección de aceitunas de almazara frente al contenido de grasa total del análisis de referencia – Variedad Picual .....	51

Figura 32 Representación de los errores de calibración (SEC) y error de validación cruzada (SECV) frente al número de longitudes de onda incluidas en el modelo – Variedad Picual .....	52
Figura 33 Izquierda) Modelo de Regresión lineal múltiple del contenido de humedad para la variedad Picual tomando 3 Longitudes de onda, (Derecha) Ídem tomado 7 longitudes de onda. ....	53
Figura 34 Izquierda: Espectros de aceitunas deshidratadas de la variedad Cornicabra sin pretratamiento, Derecha: Ídem, normalizados.....	53
Figura 35 (Izquierda) Loadings para espectros sin pretratamiento con valores elevados en torno a 1.200 y 1.400 nm; (Derecha) Ídem normalizados. - Var Cornicabra. ....	54
Figura 36 Gráfico de Cajas;( Izquierda) para el score 6 sobre las muestras crudas; (Derecha) Ídem sobre los espectros normalizados - Variedad Cornicabra.....	55
Figura 37 (Izquierda) Comparación múltiple de los datos estadísticos del ANOVA sobre el PC6 de los espectros crudos para determinar si existen diferencias significativas entre las categorías P, S y T; (Derecha) Ídem sobre los espectros normalizados – Variedad Cornicabra.....	55
Figura 38 (Izquierda) Gráfico de cajas para el análisis PLSDA con espectros crudos; (Derecha) Ídem sobresobre los espectros normalizados - Variedad Cornicabra .....	56
Figura 39 (Izquierda) Scores 1 y 2 del PLSDA de los espectros sin pretratamiento, (Derecha) Ídem en espectros normalizados – Variedad Cornicabra .....	57
Figura 40 Loadings para espectros sin pretratamiento con valores elevados en torno a 1.200 y 1.400 nm – Variedad Cornicabra.....	57
Figura 41 Gráfico de cajas del ANOVA Realizado sobre los scores del PC4 para determinar la significancia entre las categorías B y D – Variedad Cornicabra .....	58
Figura 42 Vector de proyección para el PLSDA – Variedad Cornicabra .....	59
Figura 43 Gráfico de cajas para el ANOVA realizado sobre los scores del PLSDA, para determinar la significancia entre B y D – Variedad Cornicabra.....	59
Figura 44 (Izquierda) Gráfico de cajas para el ANOVA con el PC6 y su incidencia en los días de recolección y (Derecha) Representación de la variación del contenido de grasa en los días de recolección – Variedad Cornicabra .....	60
Figura 45 Representación de los errores de calibración (SEC) y error de validación cruzada (SECV) frente al número de longitudes de onda incluidas en el modelo – Variedad Cornicabra .....	61
Figura 46 Modelo de Regresión lineal múltiple del contenido de humedad para la variedad Cornicabra. ....	62

## INDICE DE TABLAS

Tabla 1 Compuestos del aceite de oliva ((Lozano Sanchez et al., 2009) .....	20
Tabla 2 Valoración del valor de $R^2$ .....	31
Tabla 3 Procedencia de muestras de aceituna .....	36
Tabla 4 Número de espectros utilizados en los análisis para determinar los parámetros evaluados .....	40
Tabla 5 Porcentaje de grasa de cada uno de los grupos conformados en el análisis de clústeres .....	43
Tabla 6 (Izquierda) Análisis de varianza para el PC5 de los espectros crudos para determinar la significancia entre categorías; (Derecha) análisis de varianza para el PC5 de espectros normalizados para determinar la significancia entre categorías – Variedad Picual. ....	45
Tabla 7 Izquierda - ANOVA sobre los scores del PLSDA en los espectros crudos; Derecha - Ídem sobre los espectros normalizados – Variedad Picual .....	47
Tabla 8 cuadro del ANOVA Realizado sobre los scores del PC #5 – Variedad Picual .....	48
Tabla 9 Cuadro del ANOVA realizado sobre los scores del PLSDA – Variedad Picual .....	50
Tabla 10 Análisis de varianza del PC5 en función a los días de recolección para los espectros crudos de la variedad Picual.....	52
Tabla 11 Longitudes de onda seleccionadas en la regresión multilineal paso a paso para la estimación del contenido en humedad – Variedad Picual.....	52
Tabla 12 (Izquierda) Cuadro del ANOVA sobre el PC 6 con espectros no normalizados, con el factor categoría de madurez; (Derecha) Ídem sobre espectros normalizados – Variedad Cornicabra. ....	54
Tabla 13 (Izquierda) ANOVA sobre los scores del PLSDA en los espectros crudos; (Derecha) Ídem sobre los espectros normalizados – Variedad Cornicabra .....	56
Tabla 14 Cuadro del ANOVA Realizado sobre los scores del PC 4 – Variedad Cornicabra.....	58
Tabla 15 ANOVA realizado sobre los scores del PLSDA considerando como factor la presencia de defectos – Variedad Cornicabra .....	59
Tabla 16 Cuadro de ANOVA para la incidencia de los días de recolección – Variedad Cornicabra .....	60

Tabla 17 Longitudes de onda seleccionadas en la regresión multilineal paso a paso para la estimación del contenido en humedad – Variedad Cornicabra.....	61
--	----

## INDICE DE ECUACIONES

Ecuación 1 .....	24
Ecuación 2 Análisis de Componentes Principales (Otto, 2007).....	29
Ecuación 3 Modelo de regresión lineal múltiple (Rojo, 2007) .....	30
Ecuación 4 Fórmula del Coeficiente de bondad (Rojo, 2007) .....	30
Ecuación 5 Igualdad para obtener el porcentaje de humedad .....	39

## LISTA DE SIGLAS

NIR:	Espectroscopía de Infrarrojo Cercano, por sus siglas en inglés
PCA:	Análisis de Componentes Principales, por sus siglas en inglés
PLS:	Mínimos Cuadrados Parciales, por sus siglas en inglés
PLSDA:	Mínimos Cuadrados Parciales con Análisis Discriminante, por sus siglas en inglés
ANOVA:	Análisis de la varianza
SNV:	Variable normal estándar
MATLAB:	Laboratorio de Matrices

## 1 RESUMEN

En la industria oleica es ampliamente asumido que el estado de madurez, la presencia de golpes, el daño de picado de mosca u otros defectos en las aceitunas, condiciona la calidad final del aceite de oliva. La implementación de métodos baratos, rápidos y no destructivos que mejoren el conocimiento de las relaciones entre el estado de la aceituna y la calidad final del aceite es de importancia estratégica en el sector.

El presente proyecto de investigación se desarrolló en el Laboratorio de Propiedades Físicas y Técnicas Avanzadas en Agroalimentación (LPF-TAG), en el marco de una línea seguida por el grupo de investigación en colaboración con empresas del sector.

El objetivo de esta investigación es la aplicación de diferentes procedimientos de análisis de datos a espectros NIR, para la estimación de la composición de aceitunas de almazara (agua y ácidos grasos fundamentalmente) y la identificación de posibles defectos, con el fin último de establecer relaciones entre el estado de los frutos y las características del aceite de oliva obtenido.

Para el desarrollo de la investigación se utilizó las variedades Picual y Cornicabra, considerándose tres categorías de madurez: primera (P) que abarca las aceitunas verdes, segunda (S) donde se agrupan las moradas y tercera (T) con las aceitunas negras para ambas variedades, además en cada categoría se consideran también el estado de los frutos identificando cada uno de ellos como buenos (B), correspondientes a aceitunas sanas; y dañados (D), correspondientes a aceitunas con algún grado de golpe, picaduras o con cualquier otro tipo de defecto.

A lo largo de la campaña (de noviembre a febrero de 2016) se fueron tomando muestras, clasificadas según nivel de madurez y presencia o ausencia de defectos. Para cada muestra se consideró un grupo de 60 aceitunas, de las cuales se tomaron los espectros del infrarrojo cercano de forma individual en la zona ecuatorial del fruto haciéndose una media de tres mediciones por fruto.

Se aplicaron distintos preprocesados a los espectros sobre los que se realizaron diversos análisis quimiométricos. Los análisis de varianza (ANOVA) sobre las proyecciones de la población espectral en los componentes principales de los correspondientes análisis de componentes principales, mostraron diferencias significativas entre categorías de madurez para los espectros de Picual ( $F=79,50$ ) y para Cornicabra ( $F=195,29$ ) para las muestras sin normalizar; para las muestras normalizadas se obtuvo mayor significancia entre categorías ( $F=119,74$ ) para Picual, mientras que para Cornicabra el valor de  $F$  ( $F=180$ ) disminuyó, aunque las muestras siguieron siendo significativamente distintas. También se realizó el Análisis de mínimos cuadrados parciales con análisis discriminante (PLSDA) que permitió mejores valores de  $F=331,7$  para Picual y para Cornicabra  $F=569,8$ , mostrando diferencia significativa entre las categorías P, S y T para las dos variedades.

Se logró determinar diferencias significativas entre las categorías B y D a partir del PC5 ( $F=361$ ) para Picual, mientras que para Cornicabra se obtuvo  $F=645,69$  con el PC4, mejorando la significancia para el análisis PLSDA en ambas variedades ( $F=1353,94$  para la variedad Picual y  $F=1556,82$  para la variedad Cornicabra).

Para el contenido de humedad se desarrolló un modelo de regresión lineal múltiple del coeficiente de determinación de  $R^2=0,63$  para la variedad Picual y  $R^2=0,84$  para la variedad Cornicabra.

**Palabras clave:**

Espectrofotometría en el infrarrojo cercano (NIR), aceituna, aceite de oliva, ácidos grasos, humedad en el fruto, madurez, normalización.



## ABSTRACT

In the oil industry it is widely assumed that the state of maturity, the presence of shock, chopped fly damage or other defects in olives, determines the final quality of olive oil. Implementation of cheap, fast and non-destructive methods to improve knowledge of relations between the state and the final quality olive oil is of strategic importance in the sector.

This research project was developed in “*Laboratorio de Propiedades Físicas y Técnicas Avanzadas en Agroalimentación (LPF-TAG)*”, as part of a line followed by the research group in collaboration with companies in the sector. The objective of this research is the application of different methods of data analysis NIR spectra to estimate the composition of olives for oil (water and fatty acids primarily) and the identification of possible defects, with the ultimate aim of establishing relations between the state of the fruits and the characteristics of olive oil obtained.

For the development of research Picual and Cornicabra varieties used, considering three categories of maturity: first (P) covering green olives, second (S) are grouped purple olives and third (T) are grouped with black olives for both varieties, plus in each category is also the state of the fruits consider identifying each as good (B), corresponding to healthy olives; and damaged (D), corresponding to olives with some degree of stroke, bites or any other defect.

Throughout the season (November to February 2016) were taking samples were classified as level of maturity and presence or absence of defects. For each sample a group of 60 olives were considered, which the near infrared spectra were taken individually in the equatorial zone of the fruit, making an average of three measurements per fruit.

Different preprocessed were applied on the spectra, and then examined with several chemometric analyses. The analysis of variance (ANOVA) on the spectral projections of population in the main components of the corresponding principal component analysis showed significant differences between categories of maturity Picual spectra ( $F = 79.50$ ) and Cornicabra ( $F = 195.29$ ) for non-standardized samples;

They applied to different preprocessed spectra for which various chemometric analyzes were performed. The analysis of variance (ANOVA) on the spectral population projections of the main components of the corresponding principal component analysis showed significant differences between categories of maturity Picual spectra ( $F = 79.50$ ) and Cornicabra spectra ( $F = 195.29$ ) for non-standardized samples; for standard samples greater significance between categories was obtained ( $F = 119.74$ ) for Picual while for cornicabra the F value ( $F = 180$ ) decreased, although the samples were still significantly different. Analysis with partial least squares discriminant analysis (PLSDA) was asses, which allowed better values  $F = 331.7$  for Picual and  $F = 569.8$  for Cornicabra, showing significant difference between P, S and T categories for both varieties.

Was possible to determine significant differences between categories B and D from PC5 ( $F = 361$ ) for Picual, while for Cornicabra variety the F obtained was  $F = 645.69$  with PC4, improving the significance for the PLSDA analysis in both varieties ( $F = 1353.94$  for Picual and  $F = 1556.82$  for Cornicabra).

For the moisture content a model of multiple linear regression coefficient of determination with a  $R^2 = 0.63$  for Picual and  $R^2 = 0.84$  for Cornicabra variety was developed.

## 2 INTRODUCCIÓN

El cultivo del olivo, forma parte de una industria de gran importancia en España, contando actualmente con el 14% de la superficie agraria útil cultivada (MAGRAMA, 2013), lo cual lo constituye, en uno de los principales dinamizadores de la economía Española. La gama de productos comercializados por este sector, van desde aceitunas para consumo en mesa (mínima o altamente procesada) hasta la producción de aceite de oliva, siendo ambos muy demandados por el consumidor nacional y extranjero.

Es importante señalar que los estándares de calidad exigidos por el comprador son altos, lo que implica constante innovación tecnológica y revisión de los procesos por parte de la industria oleica para obtener producciones cuyas características sean óptimas para el mercado.

En este contexto, resulta importante determinar la calidad de la aceituna, en especial de la destinada a la producción de aceite de oliva, pues se sabe, que esta es la base para la obtención de un aceite de oliva de calidad. La determinación de este parámetro es compleja, pues intervienen factores como la madurez óptima, los daños en frutos, la presencia de plagas y el contenido adecuado de compuestos químicos, entre otros.

Teniendo en cuenta esta problemática, muchos esfuerzos han sido orientados a conseguir técnicas o métodos que permitan identificar la calidad de las aceitunas de forma oportuna y confiable. Destaca sobremanera la labor del grupo de investigación LPF-TAGRALIA: Técnicas Avanzadas en Agroalimentación, quienes vienen desarrollando un amplio proyecto con el fin de determinar las propiedades fisicoquímicas de la aceituna de almazara y relacionarla a la calidad del aceite de oliva, mediante técnicas de espectrometría NIR.

Es así, que el presente trabajo, forma parte del mencionado proyecto, centrándose en la aplicación de la tecnología de espectroscopia NIR en aceitunas frescas y deshidratadas para determinar los estados de madurez, contenido de humedad y contenido de grasa, mediante técnicas de análisis quimiométricos.

### 3 REVISIÓN BIBLIOGRÁFICA

#### 3.1 PRODUCCION OLEICA

La gran aceptación del aceite de oliva por sus demostradas cualidades saludables; propiedades antioxidantes (polifenoles y vitamina E), alto porcentaje de ácido graso monoinsaturado y poliinsaturados (destacando el omega 3), por lo que se conoce como el oro líquido, el cual procede del fruto del olivo, la aceituna, tras una serie de procesos extractivos.

España tiene la mayor producción de aceite de oliva a nivel mundial, representando un 60 % de producción a nivel de la Unión Europea y un 45 % a nivel mundial. Considerándose por ello el país principal de producción oleícola.

La superficie agrícola utilizada es de 2.584.564 ha de las cuales el 28 % se encuentra bajo regadío y muchas de las áreas en proceso de incorporación de riego tecnificado, logrando así mejores rendimientos con menores cantidades de agua.

El transcurso del año 2012- 2013 se incrementaron las áreas de producción como muestra la Figura 1, por otro lado la producción decayó, ver Figura 2; para la misma campaña esto quizá se deba al carácter vecero propio del cultivo.

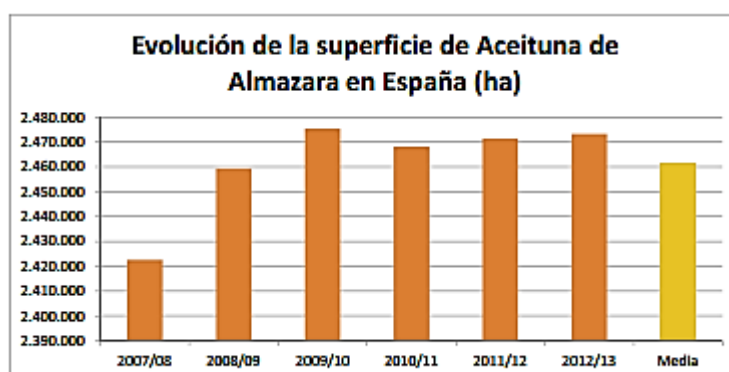


Figura 1 Evolución de la superficie de aceituna de almazara en España (ha), Fuente MAGRAMA (2013).

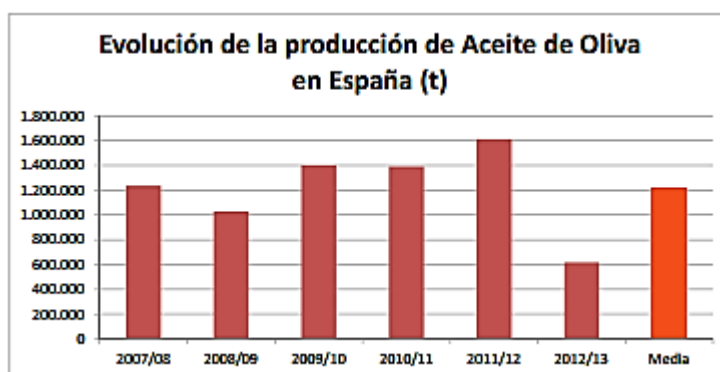


Figura 2 Evolución de la producción de aceite de oliva en España (t), Fuente MAGRAMA (2013).

La Figura 2 muestra la variabilidad de la producción durante las 6 últimas campañas, revelando una media de producción de 1.215.798 t y siendo la campaña 2011-2012 la de mayor producción.

En cuanto a la comercialización, España distribuye su producción a más de 100 países, siendo el primer exportador de aceite de oliva. Gran parte de la producción en los últimos años ha sido destinada a la exportación, siendo alrededor del 75% vendida dentro del área de la Unión Europea, salvo en la campaña 2012-2013 que fue extremadamente baja por la cantidad de producción obtenida para esta campaña. (MAGRAMA, 2013).

### 3.1.1 Descripción de la planta de olivo

El olivo pertenece a la familia botánica Oleaceae, según estudios históricos se considera que proviene de Siria, del Asia Menor, y del Oriente Próximo, sin embargo también se sabe que existían plantas silvestres en el Norte de África, España y en Grecia, por tal motivo no se conoce a ciencia cierta su origen.

El Árbol puede crecer de 4 a 8 metros dependiendo de la variedad, clima donde se ha desarrollado y tipo de poda que haya recibido el cultivo, poseen hojas simples de forma lanceolada que pueden vivir en el árbol dos o tres años, son cultivos longevos, logran vivir hasta 2000 años.

El fruto es una drupa con una sola semilla, mide aproximadamente de 1 a 4 cm y con diámetro de 0,6 a 2 cm. En madurez, la aceituna es negra, negro-violácea o rojiza, pero, en muchos casos, se cosecha antes, en estado verde. (Uceda, Hermoso, & Aguilera, 1998).

Desde la fecundación hasta la maduración transcurren de 6 a 7 meses, dependiendo de la variedad, manejo y climatología, entre otros. (Gracia & León, 2011; Picornell & Melero, 2014). Teniendo como media 200 días. La conformación del fruto completamente desarrollado lo proporciona la pulpa entre 70 y 90% del fruto, el hueso entre el 9 y 20% y la semilla entre 2 y 3%, dependiendo de los factores antes mencionados.

La producción crece en las ramas del año anterior, la asimilación y almacenamiento para la fructificación se encuentra en las hojas que pueden mantenerse 2 o 3 años en el árbol, por tal motivo el manejo debe ser constante para una buena fructificación futura y mantenimiento del cultivo. Los índices a tomar en cuenta para una buena producción son: número de ramos fructíferos, frutos por ramo, tamaño del fruto y rendimiento graso (Picornell & Melero, 2014).

### 3.1.2 Problemas sanitarios en los cultivos de olivo

El cultivo de oliva es severamente afectado por un gran número de plagas y enfermedades, las cuales menguan la calidad del fruto. A continuación se mencionan las plagas y enfermedades más importantes para la región Europea.

#### 3.1.2.1 Plagas

- La Mosca del Olivo (*Bactrocera oleae*)

Es la plaga más conocida en el cultivo de oliva y es la más persistente, es un insecto específico del olivo. Se puede encontrar en el área Mediterránea, en Asia Occidental y en diversas zonas de África.

Las pupas se encuentran en estado de latencia en invierno y las primeras moscas aparecen en primavera, cuando empiezan a cuajar los frutos depositan los huevos que al eclosionar en estado de larva ocasionan minaduras cada vez más grande a medida que van creciendo Figura 3.



*Figura 3 Izquierda: Daño de larva en fruto de oliva (Izquierda - P. del Estal Padillo, Unidad de Protección de Cultivos, E.T.S.I. Agrónomos, Universidad Politécnica de Madrid y Sección de Agricultura y Sanidad Vegetal, D.G. Agricultura y Desarrollo Rural, Comunidad de Madrid – derecha).*

Los daños en aceitunas de almazara pueden ser directos, por la disminución de la producción debida a la pérdida de peso (entre un 10% y un 15%) o a la caída prematura del fruto; pero también indirectos, por la pérdida de calidad de los aceites producidos a partir de estos frutos. En las aceitunas atacadas por la mosca del olivo se instalan fácilmente diferentes clases de hongos que desencadenan podredumbres en condiciones de humedad alta y temperatura templada, lo que origina aceites con un elevado índice de acidez y de inferior calidad organoléptica (Gomez et al., 2004).

El tratamiento es a base de insecticidas, mayormente organofosforados con atrayentes para los adultos, y pulverizaciones totales para las larvas.

- La Polilla o Prays del Olivo (*Prays oleae*)

Los adultos miden entre 6 y 13 mm de largo. Este insecto tiene tres generaciones al año que coinciden con el desarrollo del olivo, atacando hojas, flores y frutos. A inicios de junio cuando los frutos empiezan a cuajar las polillas ovipositan en el fruto por la inserción del pedúnculo, la larva come la pulpa, hasta convertirse en adulto y salir del fruto.

Los daños producidos (Figura 4) por la última generación son los más importantes, ya que son los que afectan al fruto. La primera caída de los frutos en Junio coincide con una caída fisiológica de la aceituna. Normalmente, la caída de los primeros frutos (con un tamaño de 2 a 4 mm), suele deberse a una causa fisiológica, mientras que las que caen posteriormente, ya con un tamaño de 7-8 mm, se atribuyen a la polilla (Gomez et al., 2004).



*Figura 4 Daño de polilla en fruto del olivo*

#### Daño de polilla en fruto del olivo

- La Cochinilla o Piojo Violeta del Olivo (*Parlatoria oleae*)

La cochinilla se desarrolla en las hojas y los tallos succionando la sabia del cultivo, la proliferación de la plaga puede ocasionar marchitez, sequedad y caída de hojas. En la aceituna produce unas manchas violetas (Figura 5) y deformaciones que pueden disminuir el rendimiento del aceite extraído hasta en un 20%; pero, principalmente, afecta a la calidad de la aceituna de mesa, originando pérdidas en la comercialización (Gomez et al., 2004).



*Figura 5 Incidencia de cochinilla violeta en frutos de aceituna (imagen de AgroEs.es)*

#### 3.1.2.2 Enfermedades

- La Tuberculosis

Conocida también como tumores por las malformaciones que se forman alrededor del tallo (ver Figura 6) o rama, evitando el paso de la savia hacia las partes del cultivo. La bacteria causante de la tuberculosis se denomina *Pseudomonas syringae*, en concreto una variante patogénica (pv. *Savastanoi*). Penetra en el olivo a través de heridas producidas por la poda, la recolección, el granizo o las heladas. La propagación se hace a través del agua de lluvia, por los roces de las ramas o por los instrumentos de poda.

Es una enfermedad de importancia pues su desarrollo en el cultivo produce sabor rancio y amargo en el fruto, llegando a degradar la calidad del aceite de oliva de manera significativa.



*Figura 6 Rama de olivo infestada con tuberculositis*

Es una enfermedad difícil de controlar por lo que se recomienda prevenir en el momento de podas, desinfectando el material utilizado al pasar de un árbol a otro para procurar contagiar la enfermedad entre plantas.

### 3.2 Parámetros fisicoquímicos de calidad

La calidad del aceite de oliva lo determina la composición y cuantificación de diversos compuestos químicos, los cuales se encuentran en menor y mayor proporción según la variedad del olivo, grado de maduración, condiciones agronómicas, tipo de extracción, almacenamiento y otros parámetros a tener en cuenta en el proceso de obtención del aceite de oliva. El deterioro que sufre el aceite contenido en la aceituna a causa de cualquier actuación defectuosa sobre ella, tiene dos orígenes, la hidrólisis, es decir, la reacción entre los triglicéridos del aceite y el agua, facilitada por agentes enzimáticos que dan lugar a la formación de ácidos libres, o bien, las reacciones de oxidación que rompen las cadenas de los ácidos grasos, originando alcoholes, cetonas, aldehídos que producen mal olor y sabor y que determinan el enrancia miento (Oliveras López, 2005).

El grado de acidez %, es un parámetro relacionado con los procesos hidrolíticos que sufre el aceite, esto se puede deber al almacenamiento mal ejecutado o al largo tiempo transcurrido entre la cosecha de las aceitunas y la molienda; su valor debe ser menor a 0,8% para ser que un aceite sea considerado aceite de oliva virgen extra. Cuanto menor es el grado de acidez del aceite mayor calidad tendrá el aceite de oliva virgen. Los frutos de oliva sanos tienen 0% de acidez libre, (Jiménez & Carpio, 2008; Oliveras López, 2005)

El índice de peróxidos (mEq O<sub>2</sub>/Kg), determina el estado de oxidación primario del aceite, deterioro que pueden haber sufrido los antioxidantes naturales como los polifenoles, la vitamina E y tocoferoles. Los valores permitidos son menores a 20 mEq O<sub>2</sub>/Kg.

La absorbancia en el ultravioleta (K<sub>232</sub>, K<sub>270</sub>), es un parámetro relacionado con las etapas de oxidación del aceite, iniciales y más avanzadas, respectivamente. El parámetro delta K se utiliza como criterio de pureza para detectar mezclas con aceites refinados.



### 3.2.1 Composición del aceite de oliva

El componente mayoritario de la aceituna es el agua (50%), que depende del contenido de aceite (22%). En su composición hay también azúcares (19%), por transformación del almidón inicial, celulosa (6%) y proteínas (1,6%). Otros compuestos como elementos minerales (1,6%), vitaminas ó compuestos fenólicos están presentes en menor proporción. (Oliveras López, 2005)

Tras el proceso de elaboración del aceite de oliva se reduce el contenido de agua, dividiendo su composición en dos elementos principales que se denomina fracción saponificable e insaponificable (Tabla 1). Dentro de la porción saponificable se encuentran los ácidos grasos, característicos del aceite de oliva, con un contenido en ácido oleico del 56-84% y un contenido del 321% de ácido linoleico. (Lozano Sanchez, Segura Carretero, & Fernandez Gutierrez, 2009; Oliveras López, 2005).

*Tabla 1 Compuestos del aceite de oliva (Lozano Sanchez et al., 2009)*

Composición del aceite de oliva	
Fracción Saponificable (Fracción mayoritaria)	Fracción Insaponificable (Fracción minoritaria)
Diglicéridos	Pigmentos
Mono glicéridos	Compuestos Volátiles
Ácidos grasos Libres	Poli fenoles
Triglicéridos	Tocoferoles
	Esteroles

#### - Composición Saponificable

Ácidos grasos: Los ácidos grasos presentes en el aceite de oliva son: mirístico, palmítico, palmitoleico, heptadecanoico, heptadecenoico, esteárico, oleico, linoléico, linolénico, araquídico, eicosenoico, behénico y lignocérico.

Triglicéridos: Son ésteres provenientes de la unión del trialcohol glicerina (1, 2, 3- propanotriol), con ácidos grasos. Determinadas posiciones de los ácidos grasos en las distintas posiciones del glicerol hace que se puedan presentar hasta 70 tipos de triglicéridos.

#### - Composición Insaponificable

A este grupo pertenecen las sustancias que representan el 2% de la composición del aceite de oliva y lo comprenden la clorofila, carotenoides, alcoholes, compuestos fenólicos, tocoferoles, esteroides libres y polifenoles. (Oliveras López, 2005)

### 3.2.2 Problemática con el control de calidad de fruto para la producción oleícola

La aceituna para almazara sufre daños de diversas tipologías a lo largo de su evolución en árbol, su recolección y su transporte y descarga en la industria. La presencia de plagas y enfermedades



en el cultivo son factores determinantes en la calidad del fruto, y su incidencia sobre el porcentaje de pérdidas de la producción dependerá del tipo de plaga y de su incidencia. Los impactos o aplastamientos durante la recolección, el transporte y la descarga ocasionan pérdidas de aceite y de agua, mayor susceptibilidad de los frutos a posibles contaminaciones y aceleración del deterioro postcosecha de los frutos. Todo lo cual tiene incidencia en la calidad del aceite de oliva, especialmente si el tiempo entre la recolección y la entrada en proceso sufre demoras.

El estado óptimo de madurez para la cosecha también es un factor importante a tener en cuenta, en general se considera que estados más avanzados dan lugar a mejores rendimientos en el contenido, mientras que las recolecciones en verde permiten la obtención de aceites con mejores condiciones organolépticas y nutricionales. El reto reside en poder establecer el estado óptimo de madurez con el que debe manejarse el fruto para obtener los aceites que cada industria desea obtener para satisfacer sus objetivos de calidad y rendimiento. El índice de madurez óptimo esta alrededor de los 3.5 pues se trata del intervalo en el que coinciden el porcentaje de polifenoles más elevado y el máximo rendimiento en aceite de las aceitunas. (Essiari, Zouhair, & Chimi, 2014).

### 3.3 ESPECTROMETRIA NIR

La espectrometría del Infrarrojo cercano (NIR), emplea fotones de energía para recolectar información en las longitudes de onda de 750 a 2500 nm, que pertenecen al gran espectro electromagnetico Como se puede apreciar en la Figura 7, las longitudes de onda, es decir la distancia entre los picos de onda, van creciendo a medida que se acerca a las longitudes de radio, 100km.

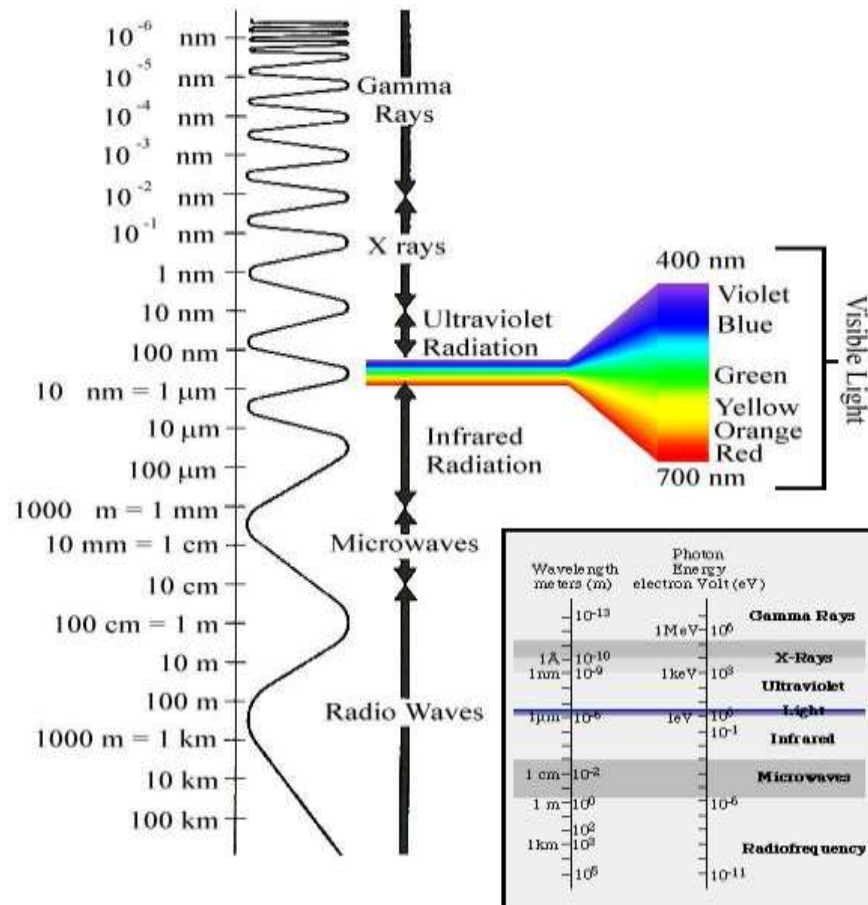


Figura 7 Arreglo del Espectro electromagnético; Fuente: Scheme of Electromagnetic spectrum (University of Prince Edward Island).

La técnica NIR se basa en la interacción de la radiación con las muestras. En la práctica las muestras son irradiadas por radiación NIR, el comportamiento de la radiación dependerá de las uniones químicas específicas de la muestra irradiada, esto se debe a que los enlaces entre los átomos vibran a una cierta frecuencia, y cada tipo de estas uniones químicas dentro de una muestra absorberá radiación NIR de una longitud de onda específica, mientras todas las demás longitudes de onda son reflejadas. (Millar, Whitworth, & Chau, 2008).

En el proceso de la toma espectral pueden darse las siguientes interacciones entre las muestras y la luz:

- Reflexión: proceso por el que la radiación electromagnética se devuelve en el límite entre dos medios.
- Absorción: es la transformación de la energía radiante a otro tipo de energía, por lo general de calor, por la interacción con la materia.
- Transmisión: paso de la radiación electromagnética a través de un medio
- Dispersión (también denominado difusión): proceso de desviar un haz de luz unidireccional en muchas direcciones.

- **Refracción:** el cambio en la dirección de una onda asociada a un cambio en su longitud de onda cuando la onda se desplaza desde un medio a otro con diferente índice de refracción. La reflexión, transmisión y dispersión salen de la longitud de onda de la radiación sin cambios, mientras que la refracción se utiliza en dispositivos espectrales para la medición selectiva de la luz a ciertas longitudes de onda (Gigahertz-Optik, 2016).

A partir de las interacciones se pueden configurar hasta tres tipos de mediciones para la toma espectral, como muestra la Figura 8:

En el caso de la reflectancia la fuente de luz y el detector están montados bajo un ángulo específico, e.g ,  $45^\circ$  , para evitar la reflexión especular, en transmitancia el modo de la fuente de luz está colocada opuesta a la del detector. Mientras que en el modo de interactancia la fuente de luz y el detector están colocado en paralelo entre sí de tal manera que la luz debido a la reflexión especular no puede entrar directamente en el detector. Este se puede lograr por medio de un cable bifurcado en el que las fibras que conduce a la fuente y el detector están en paralelo entre sí y en contacto con el producto, o por medio de un arreglo óptica especial. (Greensill & Walsh, 2000; Nicolai et al., 2007).

Cada interacción puede revelar ciertas propiedades de la materia. Con respecto a las mediciones espectrales ópticas sobre la evaluación de la calidad del fruto, reflexión y absorción son generalmente asociadas con las propiedades químicas de la fruta mientras que las propiedades de dispersión se han relacionado con firmeza de la fruta y otros rasgos físicos de la fruta. (Nicolai et al., 2007). Convirtiéndose en una técnica ampliamente usada en el análisis cuantitativo de compuestos que contengan agrupaciones funcionales con hidrógenos unidos a carbonos, nitrógenos y oxígenos ya que son la base de muchos de los principales componentes de los alimentos, tales como proteínas, hidratos de carbono, grasas y agua, la región se puede utilizar para determinar sus cantidades relativas (Millar et al., 2008).

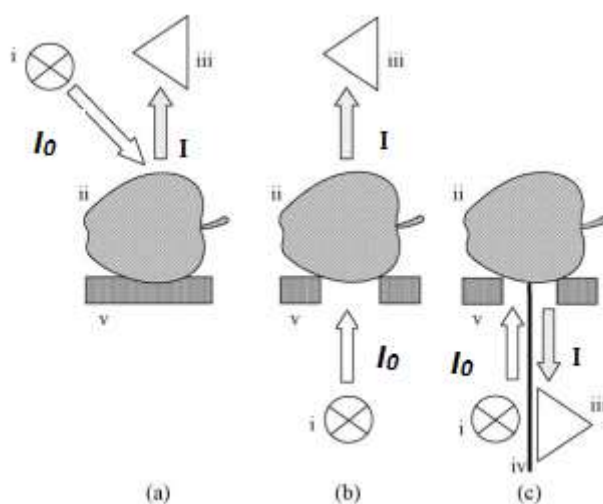


Figura 8 Configuración para la adquisición de (a) la reflectancia , ( b ) la transmitancia , y los espectros interactancia ( c ) , con ( i ) la fuente de luz , ( ii ) la fruta , ( iii ) monocromador / detector , ( iv ) la barrera de luz , y ( v ) de apoyo. En el modo interactancia, la luz debido a la reflexión está físicamente impedido de entrar en el monocromador por medio de una barrera de luz ( iv ) .  $I_0$  es la intensidad de la luz original,  $I$  es la intensidad de la luz medida por el monocromador / detector (Nicolai et al., 2007).

Por lo tanto, los espectros de reflectancia, adquirida con un sistema experimental de reflectancia, generalmente se mide tomando la relación entre reflejada (más dispersa) la luz y la luz incidente ( $R = I / I_0$ ) ; mientras que la absorbancia se define como  $A = -\log(R)$

El espectro de reflectancia aparente se ve afectada por dos fenómenos (Lleo, Roger, Herrero-Langreo, Diezma-Iglesias, & Barreiro, 2008).

- El primero es debido a las variaciones de la dispersión de la luz, que inducen variaciones en la longitud de la trayectoria media de la luz. Esto produce una variación multiplicativa de la reflectancia observada.
- La segunda es una variación aditiva de la reflectancia observada. Como primera aproximación, se considera comúnmente que el efecto multiplicativo no depende de la longitud de onda, mientras que el aditivo es una función de la longitud de onda.

Estas consideraciones producen la siguiente expresión de la reflectancia aparente:

Ecuación 1

$$R(\lambda) = K_1 R_0(\lambda) + F(\lambda)$$

Donde:

$\lambda$  es la longitud de onda,  $R_0(\lambda)$  es la verdadera reflectancia (desconocida),  $K_1$  es una constante y  $F$  una función (por lo general lineal).

La recuperación de la verdadera reflectancia de la señal medida implica dos correcciones, respectivamente relacionados con los efectos multiplicativos y aditivos:

Realizar una normalización, que consiste en dividir el espectro por un valor afectado por el efecto multiplicativo pero no ligado a la composición química de la muestra, tales como el procesamiento normal estándar variable aleatoria (SNV) lleva a cabo la normalización por la desviación estándar del espectro. (Barnes, Dhanoa, & Lister, 1993).

Otra alternativa propuesta es la sustracción de línea base, que consiste en la eliminación de la tendencia de fondo de cada espectro. Esta operación podría lograrse de forma explícita, por ejemplo mediante la identificación de la línea de base (lineal, cuadrático, etc.) o mediante la realización de una diferenciación k-orden. La última solución se realiza comúnmente mediante el algoritmo Savitsky y Golay (Gorry, 1990), realizando el suavizado de forma simultánea a la diferenciación. Como se observa por (Le Maire, Francois, & Dufrene, 2004), la idea principal es adaptarse a un polinomio diferente a cada punto de los datos en torno a este punto. Los resultados dependen de dos parámetros: el tamaño de la ventana móvil (aumento de suavizado y la disminución de la precisión) y el orden del ajuste polinómico (aumentando la precisión, pero también el ruido). Generalmente se prefiere la segunda derivada pues permite eliminar las líneas de base, como  $(a\lambda + B)$ , y magnifica los picos.

La espectroscopia NIR proporciona un conjunto de datos, más o menos complejo, que se denomina señal espectral. Dicha señal es una magnitud física que incorpora información química y/o física de la muestra analizada. Para obtener dicha información, se debe correlacionar la señal obtenida con la propiedad analítica que se pretende medir. Este proceso no es directo, ya que la señal obtenida resulta de la contribución de más de una propiedad de la muestra. Poder extraer la información útil e interpretar los datos para que puedan ser utilizados y relacionados con el parámetro a determinar es una tarea compleja, que ha dado lugar al desarrollo de los métodos quimiométricos. Así, la quimiometría permite hacer frente a diversos problemas que surgen en el desarrollo de métodos de análisis mediante espectroscopia NIR, como el filtrado de señales para eliminar o reducir el ruido, los métodos de reconocimiento de patrones, o los métodos de calibración multivariable.

### 3.3.1 Equipos espectrómetros

La región del infrarrojo cercano se descubre en el año de 1800 por Willian Herchel, pero no es sino hasta 1950 que se empiezan a construir equipos de espectrometría provistos de detectores fotoeléctricos capaces de registrar la región NIR, es a partir de entonces que se comienzan a desarrollar múltiples investigaciones teniendo como instrumento principal el espectrofotómetro. (Salguero-Chaparro, Baeten, Fernández-Pierna, & Peña-Rodríguez, 2013).

Tras el desarrollo de nuevos y mejores instrumentos se fueron creando máquinas más rápidas, fiables y con mejores prestaciones, crecía también la industria informática que hacía posible la aplicación masiva de la espectrometría NIR en diferentes ámbitos científicos. Así los espectrofotómetros fueron reemplazando métodos lentos y destructivos dentro de la tecnología analítica de productos.

Actualmente existen instrumentos NIR exclusivos de laboratorio, de los cuales las características importantes a tomar en cuenta son tamaño, peso, portabilidad, de amplio uso, que se puedan tomar muestras sólidas, líquidas, pastosas, pulverulentas, etc. También se pueden encontrar instrumentos para ser instalados en la líneas de proceso, los cuales suelen usarse en una amplia variedad de aplicaciones, tal como la supervisión y el control de los procesos de mezcla y dilución en línea, la determinación de homogeneidad y clasificación, entre otros. (Salguero-Chaparro et al., 2013).

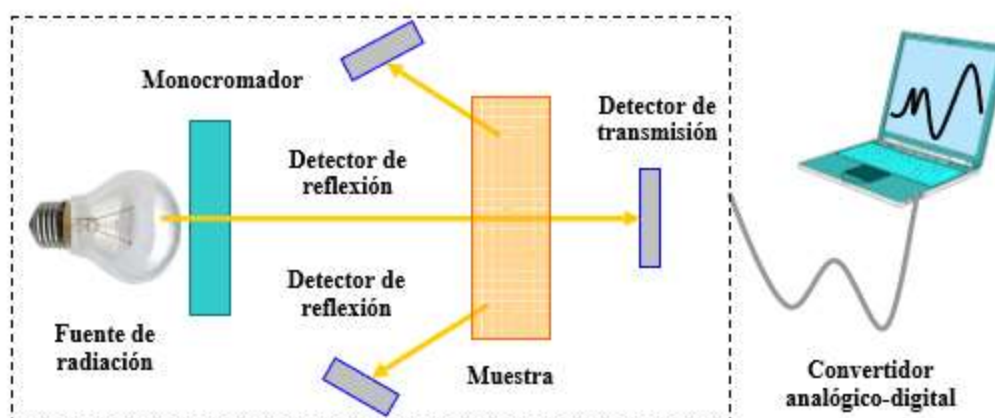
- Partes del espectrofotómetro.

Los instrumentos NIR tienen como componentes básicos la fuente de radiación (se utilizan lámparas halógenas), detector de reflexión, accesorio de presentación de la muestra, monocromador y un detector de transmisión. Como se muestra en la Figura 9.

Monocromador: Este accesorio permite clasificar a los espectrómetros. “En un instrumento de filtro, el monocromador es una rueda que sostiene un número de absorción o filtros de interferencia. Su resolución espectral es limitada. En un instrumento de escaneo, el monocromador es una rejilla o un prisma que se utiliza para separar las frecuencias individuales de la radiación ya sea de entrada o salida de la muestra. El separador de longitud de onda gira de modo que la radiación de las longitudes de onda individuales posteriormente alcanza el detector. Los espectrómetros que utilizan la Transformada de Fourier usan un interferómetro para generar luz modulada; el dominio del tiempo de la señal de luz reflejada o transmitida por la muestra se puede convertir en un espectro a través de una rápida de Fourier transformar. A menudo en estos

espectrómetros se utiliza un interferómetro de Michelson, pero también los interferómetros de polarización. En los espectrómetros de celdas de fotodiodo (PDA), una rejilla enfoca la radiación dispersa en una matriz de silicio (350-1100 nm) o un detector de fotodiodo InGaAs (arseniuro de galio indio, 1100–2500 nm). Sistemas basados en Laser no contienen monocromador, pero tienen diferentes fuentes de luz láser o un sintonizador laser. Instrumentos de filtro óptico acústico (AOTF) usan un filtro de paso de banda óptico basado difracción que puede ajustarse rápidamente para pasar a varias longitudes de onda de la luz por la variación de frecuencia de una onda que se propaga a través de la acústica por medio de cristal anisotrópico. Finalmente, los instrumentos de filtros sintonizables de cristal líquido (LCTF), estos usan un filtro birrefringente para crear interferencia constructiva o destructiva basada en el retardo, en la fase entre los rayos de luz ordinaria y extraordinaria que pasan a través de un cristal líquido. De esta manera, actúan como un filtro de interferencia de pasar una sola longitud de onda de la luz.

Mediante la combinación de varias etapas electrónicamente ajustables en serie, se puede conseguir alta resolución espectral (Nicolai et al., 2007).



*Figura 9 Componentes básicos de un instrumento NIRS; (Salguero-Chaparro et al., 2013)*

### 3.3.2 Aplicaciones

La técnica NIR se utiliza en la agroindustria, en la industria farmacéutica y petroquímica obteniéndose resultados óptimos, en muchos casos se presenta como análisis de garantía de calidad, mostrando una clara expansión mediante distintas investigaciones. Además por todas las ventajas que simboliza esta técnica es posible que en un futuro reemplace otras técnicas (Abbott, 1999; Alcalà Bernàrdez, 2006; Bautista Mercader & Blanco Romia, 2009).

En el sector agronómico se han realizados numerosas investigaciones buscando explorar las propiedades de distintos productos aplicando espectrometría NIR, así Gracia y León (2011) evaluaron la determinación de la humedad y el rendimiento graso en aceitunas mediante un espectrómetro portátil, obteniendo predicciones aceptables. Se ha extendido el uso de Cámaras digitales con filtros de Bandas del infrarrojo cercano para distintos fines, Guzmán et al., (2013)

evalúa el daño externo en aceitunas de mesa y almazara, utilizando esta técnica, y logra determinar que es una tecnología útil para la automatizar la evaluación de las aceitunas en las líneas de producción. De otra parte se Evaluaron aceitunas en las líneas de producción, en movimiento, con un instrumento NIR de celdas de diodo, con el objetivo de determinar la humedad, el contenido de grasa y acidez concluyendo que los resultados de predicción de NIR en línea de producción fueron aceptables para los tres parámetros. (Salguero-Chaparro et al., 2013).

Por otro lado, dentro de la industria farmacéutica se está implementando la espectrometría NIR como una herramienta importante para la Tecnología Analítica de Procesos. Estudiándose un proceso de monitorización on-line en procesos de granulación en una planta de producción mediante el uso de una sonda NIR, con resultados alentadores. Y por ser una alternativa rápida y eficiente se convierte en una alternativa a los métodos convencionales. (Bautista Mercader & Blanco Romia, 2009).

### 3.4 METODOS QUIMIOMETRICOS

#### 3.4.1 Quimiometría

Con la aparición de nuevas herramientas de automatización e informática, capaces de ayudar a obtener grandes cantidades de datos en el laboratorio es que aproximadamente hace 30 años se establece el término de quimiometría cuyo objetivo final es interpretar grandes cantidades de data para convertirla en información, mediante métodos matemáticos, estadísticos y otros . En términos generales, la quimiometría es una expresión que resume el uso de métodos matemáticos, estadísticos y otros. (Grupo de Quimiometría y Cualimetría de Tarragona, 2002)

La quimiometría permite analizar gran cantidad de datos, abarcando hasta tres etapas en el proceso de análisis, la primera es el pre tratamiento de la data, con el cual logramos mejorar su calidad; La segunda es la etapa de reconocimiento de pautas, a través de diferentes técnicas multivariantes que permiten discriminar agrupaciones de muestras en base a sus similitudes, finalmente la tercera etapa corresponde a los métodos de calibración, que permiten establecer una relación cuantitativa entre la señal analítica y las propiedades de la muestra. (Salguero-Chaparro et al., 2013). Basándose en la capacidad de combinar adecuadamente las variables medidas para formar otras nuevas, denominadas variables latentes o factores, que contengan la información original. (Grupo de Quimiometría y Cualimetría de Tarragona, 2002).

Debido al gran número de herramientas de análisis útiles se convierte en una solución cuando hablamos de análisis multivariable, así para los análisis de espectrometría en el infrarrojo cercano (NIR) donde al poseer muchas variables se hace más complejo el análisis e interpretación de datos, algunas de las técnicas matemáticas más comunes son el análisis de componentes principales (PCA), los mínimos cuadrados parciales (PLS), la regresión por componentes principales (PCR) o la regresión multilínea (MLR).

#### 3.4.2 Pretratamiento espectral

La toma espectral puede estar alterada por diversos factores externos, en este caso se procede con el pretratamiento de los datos para poder eliminar ruido y efectos que puedan influir de forma negativa al análisis de la información de la muestra. (Ver el apartado 3.3, efecto multiplicativo y aditivo). Esta modificación de datos debe ser realizada antes de construir un modelo matemático



o antes de aplicar otra forma de analizar los datos. En términos generales, el pretratamiento permite eliminar la varianza producida por errores sistemáticos en la experimentación.

Para el presente proyecto se empleó el procedimiento de normalización “*Standard Normal Variate*” (SNV), sin embargo se mencionan también otras técnicas ampliamente utilizadas en el pretratamiento de datos espectrales.

- **Suavizado:** Existen métodos que permiten disminuir el ruido mediante cálculos polinómicos de convolución (*Savitzky-Golay*) o por transformada de Fourier.
- **Derivadas:** La primera derivada elimina los desplazamientos de línea base, mientras que la segunda derivada elimina los términos que varían linealmente con la longitud de onda. lo cual permite aumentar las diferencias entre las bandas anchas y solapadas del espectro.
- **Normalización:** La normalización intenta corregir los incrementos y disminuciones que sufren las medidas de las variables, mediante la identificación de algún aspecto de cada muestra que debería ser esencialmente constante de una muestra a otra, y realizar la corrección de la escala de todas las variables utilizando dicha característica. Se normaliza cuando deseamos saber si existe relación entre las variables de los datos analizados y no cuando nuestra prioridad es saber la magnitud absoluta de la respuesta.
- **Variable normal estándar (SNV):** Se utiliza mayormente para reducir la dispersión producida por diferencias en el tamaño de partícula de muestras sólidas. De forma general se normaliza restando a cada espectro su media y dividiendo por la desviación típica.

### 3.4.3 Métodos de Reconocimiento de Pautas

Permiten la identificación de los productos químicos implicados en la investigación. Los Métodos de Reconocimiento de Pautas (PRM, Pattern Recognition Methods) son un conjunto de técnicas quimiométricas que permiten realizar análisis cualitativos de un modo rápido y simple, los PRMS se pueden clasificar de un modo general según sean supervisados o no-supervisados.

#### 3.4.3.1 Métodos no supervisados

Los métodos clasificados en este grupo tienen como finalidad ver las semejanzas y diferencias entre los espectros de las muestras, desconociendo la clase a la cual pertenecen las muestras analizadas.

- **Análisis de Componentes Principales (PCA)**

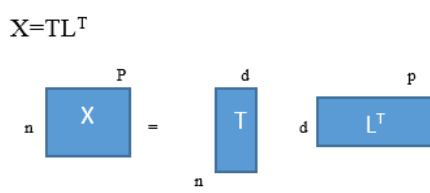
Si tenemos una amplia cantidad de variables es probable que algunas midan lo mismo pero bajo diferentes puntos de vista, por lo que se acumulan datos sin valor adicional, y además el análisis de estos se hace más complejo, siendo difícil visualizar relaciones entre las variables. Por este motivo, el uso del PCA resulta ser de gran ayuda al ser una técnica que permite reducir el número de variables y mostrarlas en dos o tres dimensiones y observar cuán variables o similares son entre sí, con la salvedad de no reducir la información ni perderla.



Las variables obtenidas, llamadas también conjunto de componentes principales son combinaciones lineales de las anteriores, que están incorreladas y se van construyendo según el orden de importancia en cuanto a la variabilidad total que recogen de la muestra.

Para visualizar las relaciones entre los espectros de las muestras se representan los scores de un determinado  $PC_1$  contra el scores de otro  $PC_2$ . En el gráfico se podrá visualizar una agrupación entre los puntos que representen las muestras donde exista relación y de lo contrario se observará una dispersión entre las muestras que no tengan semejanza.

La idea principal del PCA es aproximar la matriz original "X" con el producto de dos matrices pequeñas, llamadas score y loadings. Los loadings son los cosenos de los ángulos que forman los nuevos ejes con los originales y los scores son las coordenadas de las muestras en los nuevos ejes (Otto, 2007).

$$X = TL^T$$


The diagram shows the matrix equation  $X = TL^T$ . Matrix  $X$  is represented as a blue rectangle with dimensions  $n$  (rows) and  $p$  (columns). Matrix  $T$  is a blue rectangle with dimensions  $n$  (rows) and  $d$  (columns). Matrix  $L^T$  is a blue rectangle with dimensions  $d$  (rows) and  $p$  (columns). The equation is shown as  $X = T L^T$  with the matrices represented by these rectangles.

*Ecuación 2 Análisis de Componentes Principales (Otto, 2007)*

Donde

X: Datos originales de la Matriz, con n filas y p columnas.

T: matriz de los scores con n filas y d columnas (número principal de componentes PCs)

L: Matriz de los loadings con d columnas y p filas y T elevada es la transpuesta de la matriz.

#### - Análisis de agrupamientos (clusters)

El análisis cluster es un conjunto de técnicas multivariantes utilizadas para clasificar a un conjunto de individuos en grupos homogéneos.

Pertenece al grupo de análisis no supervisados por que los grupos son desconocidos a priori y permite un análisis de marcado carácter exploratorio, el cual nos permite agrupar a los individuos que presenten características muy similares y que a la vez se diferencien de otros grupos.

Estos grupos son fáciles de representar a partir de un dendrograma, donde de forma jerárquica se muestra cómo los grupos grandes se van dividiendo en grupos más pequeños haciendo particiones sucesivas a "distintos niveles de agregación o agrupamiento".

### 3.4.3.2 Métodos supervisados

Se trata de análisis orientados a evaluar datos de forma controlada y se caracterizan principalmente por conocer la clase a la cual corresponde la muestra a analizar, antes de realizar el propio análisis.

#### - Regresión lineal múltiple

En el caso de modelado multivariante, pueden operarse muchas variables dependientes como independientes (Otto, 2007). El modelo de regresión lineal múltiple es idéntico al modelo de regresión lineal simple, con la única diferencia de que aparecen más variables explicativas:

*Ecuación 3 Modelo de regresión lineal múltiple (Rojo, 2007)*

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \dots + b_k \cdot x_k + u$$

Al igual que en regresión lineal simple, los coeficientes  $b$  van a indicar el incremento de la variable dependiente por el incremento unitario de la correspondiente variable explicativa. Por lo tanto, estos coeficientes van a tener las correspondientes unidades de medida. (Rojo, 2007)

La  $R^2$  es el coeficiente (estadístico) que mide la bondad del ajuste del modelo. El cual se representa mediante la proporción de la varianza explicada (VE) entre la varianza total (VT); de este modo, definimos el coeficiente de determinación  $R^2$ :

*Ecuación 4 Fórmula del Coeficiente de bondad (Rojo, 2007)*

$$R^2 = \frac{VE}{VT} = \frac{VT - VNE}{VT} = 1 - \frac{VNE}{VT}$$

Donde

$R^2$ : Coeficiente de determinación

VE: Varianza explicada

VT: Varianza total

VNE: Varianza no explicada

Este coeficiente es muy importante pues determina qué porcentaje (en tanto por uno) de la varianza de la variable dependiente es explicado por el modelo de regresión. Cuanto más se acerque el valor de  $R^2$  a la unidad mejor será el ajuste entre la predicción del modelo y los valores. Y se puede clasificar los valores de  $R^2$  de la siguiente manera:

Tabla 2 Valoración del valor de  $R^2$ 

Rango de Valores de $R^2$	Menor de 0,3	0,3 a 0,4	0,4 a 0,5	0,5 a 0,85	Mayor a 0,85
Clasificación	Muy malo	Malo	Regular	Bueno	Sospechoso

Fuente: Rojo (2007).

#### - Regresión de mínimos cuadrados parciales (PLS)

La regresión de mínimos cuadrados parciales (PLS) es una técnica que reduce los predictores a un conjunto más pequeño de componentes no correlacionados y realiza una regresión de mínimos cuadrados con estos componentes, en lugar de hacerlo con los datos originales. La regresión de PLS resulta especialmente útil cuando se tiene más predictores que observaciones y la regresión de mínimos cuadrados ordinarios produce coeficientes con altos errores estándar o falla por completo.

El objetivo de esta técnica de regresión es maximizar la covariancia que existe entre un conjunto de datos  $X$  y sus correspondientes clases  $Y$ .

Este método realiza una reducción de la dimensión, generando nuevos ejes de proyección llamados variables latentes, las cuales son necesarias para representar la mayor cantidad de variación y construir un modelo de respuesta satisfactorio.

#### - Mínimos Cuadrados Parciales con Análisis Discriminante (PLSDA)

Los objetivos del análisis discriminante son encontrar funciones lineales que mejor discriminen dos grupos y construir una regla de asignación para un objeto nuevo. (Medina, 2013). PLS también se puede utilizar para el análisis discriminante o de clasificación. El enfoque más simple es cuando sólo hay dos clases. En tal caso, se colocan un vector ' $c$ ', que consiste en un número de acuerdo a qué clase pertenece un objeto, hay dos variaciones comunes, utilizando '1' y '0' o '1' y '-1' para las clases A y B. Y también

PLS puede ser utilizado para formar un modelo de calibración entre el bloque de ' $x$ ' que puede, por ejemplo, ser un conjunto de mediciones espectroscópicas, y el bloque de ' $c$ ' que puede proporcionar información en cuanto a los orígenes biológicos de una muestra. Los out put obtenidos por el PLSDA son los score, loading y pesos que se pueden utilizar para obtener información de las categorías evaluadas. Permitiendo así describir el conjunto de variables explicativas y la predicción de las respuestas. (Brereton, 2007).

## 4 PLANTEAMIENTO DEL PROBLEMA

En los últimos años en la industria oleícola se ha elevado la demanda de producción como consecuencia de la gran popularidad del aceite español en el mercado externo, constituyéndose como una industria de altas exigencias en los estándares de calidad; existen muchos factores a tomar en cuenta para la elaboración del aceite de oliva de calidad, así por ejemplo la presencia de daños en la piel del fruto es el factor más decisivo para determinar la calidad externa del fruto (Guzmán et al., 2013).

Desde el punto de vista industrial las tres cualidades más importantes de los frutos que definen la calidad del aceite que se obtiene, son el contenido de humedad de la fruta, la acidez libre y el contenido de aceite (Salguero-Chaparro et al., 2013), estos parámetros son de mucho interés para establecer el precio del producto y desarrollar de forma óptima los procesos de extracción (Delgadillo, Nunes, & Barros, 2012). El uso del espectrómetro NIR portable permite estimar el contenido de aceite y humedad en frutos de oliva medidos directamente en el árbol o en condiciones de laboratorio. Evaluando varios parámetros de forma no destructiva y en una sola medición. (Gracia & León, 2011).

Mediante el desarrollo de modelos cualitativos o cuantitativos y métodos quimiométricos basados en datos obtenidos por espectrometría NIR se podría lograr predecir la calidad de la producción oleícola, sabiendo que los espectros que caracterizan las bandas de absorción del agua están alrededor de 1460nm y 1950nm, y alrededor de 1210nm y 1730nm para el aceite. (Gracia & León, 2011).

## 5 ANTECEDENTES Y OBJETIVOS

### 5.1 Antecedentes

La espectroscopia NIR se basa en el análisis de la interacción de la radiación electromagnética en el rango de longitudes de onda comprendido entre 780 a 2500 nm. En comparación con otras técnicas no destructivas, la espectroscopia NIR presenta las ventajas genéricas de no necesitar ninguna preparación de la muestra y la posibilidad de un sencillo y rápido análisis simultáneo de varios compuestos presentes en las muestras. En la actualidad las técnicas espectrofotométricas son ampliamente empleadas en el análisis de alimentos, llegando a algunas aplicaciones en línea para la supervisión y control de la producción de alimentos. (Armenta, Moros, Garrigues, & Guardia, 2010).

Si bien existe una amplia gama de equipos espectrofotométricos comerciales con prestaciones y precios diversos, los modelos robustos disponibles para la estimación de parámetros de calidad o composicionales no son tan numerosos, y se circunscriben principalmente a productos que posibilitan la homogeneidad en la presentación de la muestra y que presentan un gran interés económico. Queda por tanto un amplio campo para el desarrollo de modelos de estimación robustos y transferibles de parámetros de calidad en productos como los hortofrutícolas.

En el sector oleícola existe un interés creciente, recogido en el Plan de Acción para el sector del aceite de oliva, en el desarrollo de técnicas rápidas y en lo posible no destructivas para garantizar la autenticidad y calidad del aceite de oliva. Actualmente existen métodos alternativos a los especificados en los reglamentos. Entre estos métodos se incluye la espectroscopia en el rango visible y NIR (Aparicio, Morales, Aparicio-Ruiz, Tena, & García-González, 2013; Frankel, 2010). Dando un paso más allá, establecer relaciones entre la composición y estado de las aceitunas y las características de los aceites obtenidos a partir de las mismas, sería un avance considerable para los productores de aceite de oliva.

En la industria del aceite de oliva, se ha aplicado la espectroscopia NIR en la determinación de composición en pulpa y pasta de aceituna (Armenta et al., 2010), o en fruto entero (Cayuela JA, Garcia JM, Caliani N., 2009), incluso en línea (Salguero-Chaparro et al., 2013 y 2014). Se ha logrado estimaciones aceptables para el contenido graso, la humedad y la acidez libre. En diversas investigaciones recomienda evaluar las aceitunas en diferentes estados de madurez para obtener resultados más amplios de los parámetros analizados. En otras líneas de investigación centradas en la mejora genética, se han aplicado equipos portátiles NIR para estimar el contenido de aceite y la humedad en aceitunas intactas tanto en el árbol (Gracia & León, 2011)

El éxito de estas investigaciones depende en buena medida de la metodología seguida para el pre-tratamiento de las bases de datos espectrales (suavizado, promediado, corrección de línea base...) y para el desarrollo de los modelos (métodos supervisados o no supervisados, con reducción de variables, de estimación, de clasificación...) (Nicolai et al., 2007).

En el marco de un proyecto de colaboración con empresas del sector, el LPF-TAGRALIA está abordando la aplicación de diversas técnicas ópticas y de imagen para la evaluación de la calidad y el estado de madurez de aceituna intacta a la entrada de la almazara. Se ha planteado un diseño experimental que aborda la toma de datos a lo largo de toda una campaña en una comarca de Toledo considerando las variedades más representativas de la zona, Picual y Cornicabra. Dentro

de este trabajo la espectroscopia NIR es una de las técnicas aplicadas. El presente TFM contiene parte de las actividades y resultados relacionadas con esta técnica en el citado proyecto.

## 5.2 Objetivos

El objetivo general que se establece en el presente TFM es explorar la viabilidad de establecer relaciones entre el estado de madurez y la presencia o no de defectos en la superficie de aceitunas intactas con sus correspondientes espectros NIR.

Para ello se definen dos objetivos específicos:

- Estimar la composición de aceitunas de almazara (agua y ácidos grasos fundamentalmente) con la aplicación de la técnica de espectrometría NIR.
- Evaluar la influencia de presencia de defectos en la superficie de aceitunas intactas en el patrón espectral NIR.

## 6 METODOLOGÍA EXPERIMENTAL

### 6.1 Colectivo Muestral

Las muestras de aceituna se recolectaron en distintas fincas de la provincia de Toledo como muestra la Tabla 3. Se evaluaron las variedades Cornicabra y Picual. La clasificación de las muestras fue realizada mediante un equipo de visión multiespectral instalado en línea, MULTISCAN S30B, calibrado para seleccionar tres categorías (de una mezcla de aceitunas como se aprecia en la Figura 10) como se indica a continuación, ver Figura 11.

- Primera (P): grupo que abarca las aceitunas verdes y pintonas
- Segunda (S): grupo que abarca las aceitunas moradas
- Tercera (T): grupo que abarca las aceitunas negras

Cada categoría se vuelve a subdividir en dos grupos, denominándolas

- Buenas (B): agrupa aceitunas sanas
- Con Defectos (D): agrupa las aceitunas con algún golpe, picadura o con cualquier otro tipo de defecto.

Para cada categoría se utilizaron un grupo de 60 aceitunas, de las cuales se tomaron los espectros del infrarrojo cercano de forma individual en la zona ecuatorial del fruto haciéndose una media de tres repeticiones por fruto.



*Figura 10 Mezcla inicial de aceitunas por categorizar*



*Figura 11 Categorías de aceitunas; verde y pintonas (P), moradas (S) y negras (T) de izquierda a derecha*

Una vez recolectadas las muestras son llevadas al laboratorio, donde la selección automática se apuraba a mano para seleccionar por color los 60 frutos de cada muestra, y asegurarse de que las muestras con defectos contuviesen todas las aceitunas afectadas por algún problema en su superficie. (Ver Figura 12).



*Figura 12 Colectivo muestral de aceitunas*

La recolección de muestras comprendió del 05 de noviembre del 2015 al 17 de febrero del 2016, teniéndose diferentes días de recolección por variedades y lugar de procedencia, lo que se resume en la Tabla 3.

*Tabla 3 Procedencia de muestras de aceituna*

Fecha	Variedad	Procedencia
05-nov-15	Picual	Zona 1
10-nov-15	Picual P1	Zona 1
	Picual P2	Zona 1
11-nov-15	Picual	Zona 1



16-nov-15	Cornicabra	Zona 1
17-nov-15	Cornicabra	Zona 1
23-nov-15	Picual	Zona 1
24-nov-15	Picual	Zona 1
	Cornicabra	Zona 2
01-dic-15	Picual	Zona 1
02-dic-15	Cornicabra	Zona 2
21-dic-15	Cornicabra P1	Zona 3
	Cornicabra P2	Zona 1
12-ene-16	Cornicabra	Zona 1
13-ene-16	Cornicabra	Zona 1
19-ene-16	Cornicabra	Zona 1
20-ene-16	Cornicabra	Zona 1
25-ene-16	Cornicabra	Zona 4
26-ene-16	Cornicabra P1	Zona 5
	Cornicabra P2	Zona 1
01-feb-16	Cornicabra	Zona 5
02-feb-16	Cornicabra	Zona 4
17-feb-16	Cornicabra	Zona 6

## 6.2 Instrumentación

Para la investigación se utilizó el espectrofotómetro Hamamatsu photonic multichannel analyser NIR (ver Figura 13), perteneciente a la serie PMA 11, el cual consiste en un espectrógrafo compacto, un foto detector multicanal, una sonda de fibra óptica aleatorizada y el circuito de control. Por su tamaño se le considera un instrumento de mesa, tiene un rango de longitud de onda de 896nm a 1684 nm.

El espectrógrafo y el detector multicanal están mecánicamente fijados. (HAMAMATSU PHOTONICS, 2003).



*Figura 13 Equipamiento electrónico del espectrómetro.*

Otros instrumentos que también se utilizaron fueron un horno secador, balanza analítica, Toda la experimentación se llevó a cabo en las instalaciones de grupo de investigación LPF-TAGRALIA de la Universidad Politécnica de Madrid.

### 6.3 Software

El software utilizado para la gestión de datos de la toma espectral fue el programa PMA U6039-01. Se selecciona la opción reflectancia, se establecen 3 repeticiones, se hacen las medias sobre esas tres medidas.

Por otro lado se empleó el programa MATLAB para el procesamiento de datos, Matlab (The Mathworks, Inc.) es el nombre abreviado de “MATrix LABoratory”. Es un programa de cálculo técnico y científico, que permite realizar operaciones con vectores y matrices. Tiene su propio lenguaje de programación y dispone de un código básico y de varias librerías especializadas (toolboxes). (Salguero-Chaparro et al., 2013).

### 6.4 Métodos y análisis de referencia

Se enviaron a un laboratorio certificado lotes representativos de las muestras, a los cuales se realizaron diferentes tipos de análisis, como acidez libre, porcentaje de grasas, contenidos en agua, rendimiento graso, contenido de aceites oleico, palmítico, linoleico; estas determinaciones constituyen, aunque no son hechas directamente sobre los mismo frutos sobre los que se toman los espectros NIR, son las referencias que pueden ser consideradas en el presente trabajo. Hay que añadir también, que estas determinaciones requieren de una mínima cantidad de aceitunas para realizar la extracción del aceite sobre el que se analizan los parámetros indicados, por lo que se trata de valores de referencia medios y no de los correspondientes fruto a fruto.

Para el análisis de muestras se utilizaron dos técnicas exploratorias, una supervisada y otra no supervisada como son el PCA y el PLSDA respectivamente, estas evaluaciones se hicieron sobre los espectros crudos y normalizados. Para predecir la humedad se hicieron análisis de regresión multilineal a las muestras frescas de cada variedad.

#### 6.4.1 Toma espectral

Las medidas espectrales se realizaron en condiciones de laboratorio, apuntando la sonda de fibra óptica hacia el diámetro ecuatorial del fruto de aceituna de forma individual y cliqueando sobre el botón de medida del programa. (Ver Figura 14). Este procedimiento se realiza para las aceitunas frescas y para las deshidratadas, tomándose 3 repeticiones por aceituna.

Para que la calidad del espectro sea bueno la intensidad debería estar entre 20000 y 30000 niveles de intensidad, a lo largo de la curva.



Figura 14 (Izquierda) Toma espectral y vista del espectrofotómetro, (Derecha) acercamiento de la toma espectral.

#### 6.4.2 Determinación del contenido en humedad

La determinación de humedad de la muestra se basa en la pérdida de humedad para lo que se llevan las muestras al horno por 24 horas con temperatura controlada a 103°C como se muestra en la Ecuación 5. Se retiran las muestras del horno y pesan nuevamente, pero esta vez deshidratadas.

*Ecuación 5 Igualdad para obtener el porcentaje de humedad*

$$H = \frac{(P_o - P_1) * 100}{P_o}$$

Donde:

H: Porcentaje de humedad

P<sub>o</sub>: Peso de la muestra fresca

P<sub>1</sub>: Peso de la muestra deshidratada

Posterior al pesado se procede a la toma espectral de las muestras deshidratadas y a guardar los datos en formato tpx. , para posteriormente procesarlos en MATLAB.



*Figura 15 Muestras secadas en horno*

#### 6.4.3 Análisis de datos

Se importaron los datos en formato tpx. al programa MATLAB versión 2015 para los posteriores análisis quimiométricos. Los análisis se realizaron de forma separada, por variedades. El análisis de varianza se ejecutó al 95% de confiabilidad para todos los casos.

Los análisis se realizaron sobre tomas espectrales sin pre tratamiento, y datos normalizados, para las categorías de madurez P, S y T, el método de normalización utilizado fue el Variable normal estándar (SNV), donde se normaliza restando a cada espectro su media y dividiendo por la desviación típica, para reducir la dispersión de los datos. Para la determinación de los parámetros de calidad (B y D); humedad y rendimiento graso se analizaron los espectros sin normalizar.

Para evaluar las categorías P, S y T se realizaron PCA sobre los espectros no normalizados y normalizados (según snv) y PLSDA sobre los scores de los primeros 10 componentes principales obtenidos en el PCA correspondiente. Se realizó también PCA y PLSDA para la determinación de los parámetros B y D sobre los espectros tomados en aceitunas frescas (n=3043 para Cornicabra y n=1353 para Picual). Para la estimación del contenido en humedad se aplicó una regresión multilíneal a los espectros tomados sobre las aceitunas frescas y buenas (B), con un total de 1824 espectros de la variedad Cornicabra y 798 de Picual. El Rendimiento graso se determinó sobre los scores obtenidos de los espectros sin normalizar, aceitunas deshidratadas; n=2851 para Cornicabra y n=1127 de Picual, como se muestra en la Tabla 4.

*Tabla 4 Número de espectros utilizados en los análisis para determinar los parámetros evaluados*

Parametros a evaluar	Número de Espectros Usados		
	Variedades		Condición de la aceituna: Frescas/ Deshidratadas
	Picual	Cornicabra	
Categorías P,S y T	1127	2851	Deshidratadas

Categorías B y D	1353	3043	Frescas
Humedad	789	1824	Frescas - Buenas
Contenido Graso	1127	2851	Deshidratadas

Es conveniente mencionar que en las figuras aparece una categoría aparte de las mencionadas, esta hace referencia a una categoría que agrupa una mezcla de frutos de aceituna con una alta heterogeneidad en cuanto a sus características de madurez y calidad (buenas y malas), se trata de una muestra de aceitunas tomada directamente del remolque que llega a la almazara, sin someterla a ninguna clasificación ni por madurez ni por defectos.

## 7 RESULTADOS Y DISCUSIÓN

Los resultados están en base a los análisis realizados sobre un total de 8374 espectros, correspondientes a las variedades Cornicabra (n= 5894) y Picual (2480). Como se muestra en la Tabla 4.

### 7.1 Variedad Picual

#### *Análisis preliminares*

Durante las primeras fechas de recolección (hasta el 21 de diciembre) se llevaron a cabo unos análisis preliminares sobre los espectros NIR de las muestras manejadas hasta aquel momento. Se aplicó análisis de componentes principales (PCA) sobre los espectros originales de los frutos sin defectos, adicionalmente se aplicó un análisis de cluster no supervisado sobre los ‘scores’ que se muestran más relacionados con el contenido de grasa y de humedad.

En la siguiente figura se presentan los pesos (‘loads’) del PCA aplicado sobre los espectros de los frutos sin defectos. Se observa que es el PC4 el más relacionado con el contenido en grasa (con un pico de absorción en torno a 1.200 nm).

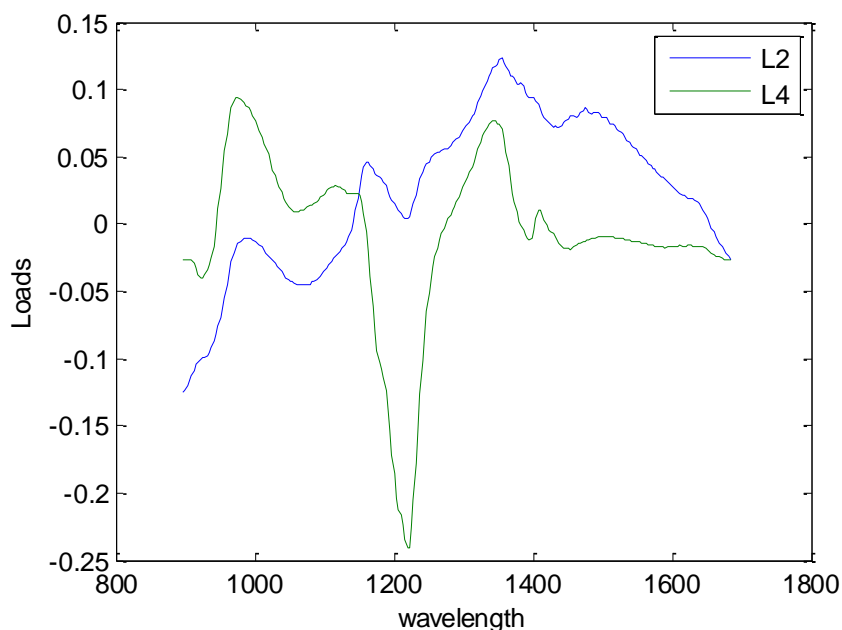


Figura 16 Loading correspondiente a la aplicación de ACP sobre los espectros no tratados de aceitunas frescas cosechadas hasta el 21 de diciembre.

La proyección de los espectros (tanto de la muestra en fresco como deshidratada) en el espacio de los componentes principales PC2 y PC4, permite diferenciar dos grupos sin solape a lo largo del eje del PC2: las aceitunas frescas (a la izquierda) y las aceitunas secas (a la derecha); indicando la relación entre el contenido en humedad y este componente principal. La aplicación de un análisis de conglomerados no supervisado sobre los scores de los frutos deshidratados da lugar a tres grupos (indicados en la leyenda como 1, 2 y 3), que se cree relacionado con el contenido graso. Para comprobar esta relación, una selección de aceitunas pertenecientes a cada clúster se envió al laboratorio para obtener el valor de contenido graso medio de cada clúster. En la Tabla

X se incluyen los resultados, en los que se observa que el menor contenido graso corresponde al clúster 3 y el mayor al clúster 2, ordenándose según valores decrecientes del PC4. Dado que la capacidad de segregación entre contenido graso es mayor en las muestras deshidratadas, la mayor parte de los análisis que se incluyen en los siguientes apartados se realiza sobre las muestras deshidratadas.

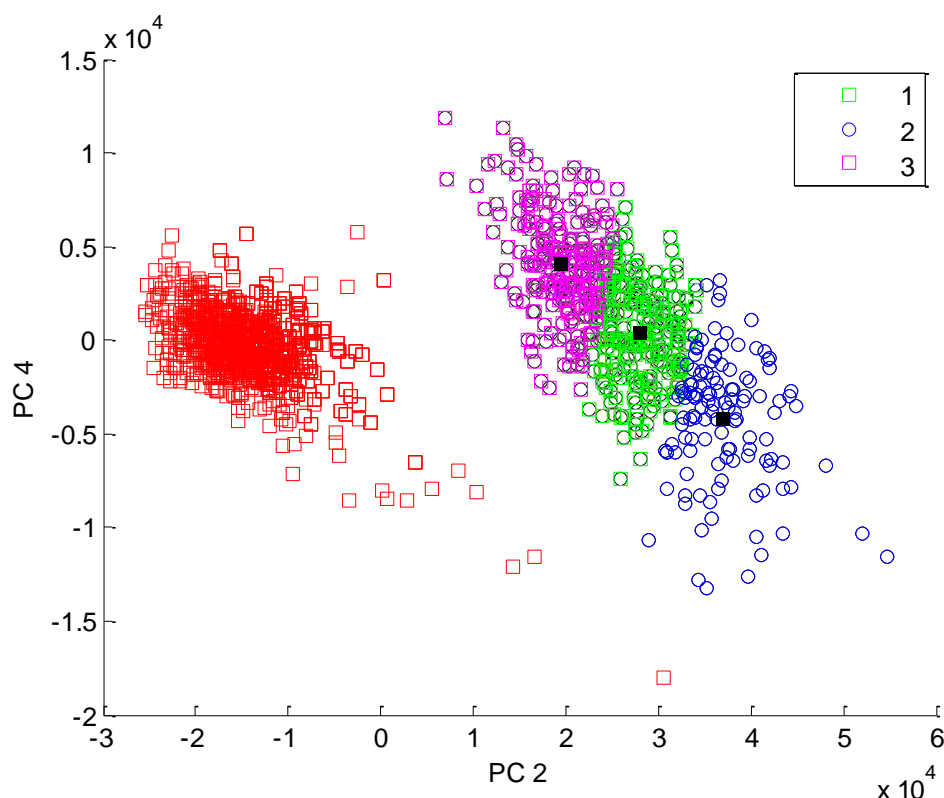


Figura 17 Proyección de las aceitunas frescas (cuadrados rojos) y deshidratadas (resto de puntos), categorizados según la agrupación del análisis de clústeres.

Tabla 5 Porcentaje de grasa de cada uno de los grupos conformados en el análisis de clústeres.

Clúster	Porcentaje de grasa
1	20,6
2	22,8
3	19,2

### ***Análisis sobre el conjunto de muestras***

A continuación se presentan los resultados de la variedad **Picual**,

La Figura 18 (Izquierda) representa los espectros crudos de la variedad Picual en las longitudes de onda de la región NIR, se puede apreciar que alrededor de la región de los 1200 nm existe un valle pronunciado también alrededor de los 1400nm, lo que se aprecia mejor en los espectros normalizados, Figura 18 (Derecha). Se sabe que esta zona representa las bandas de absorción de grasa y agua respectivamente (Gracia & León, 2011).

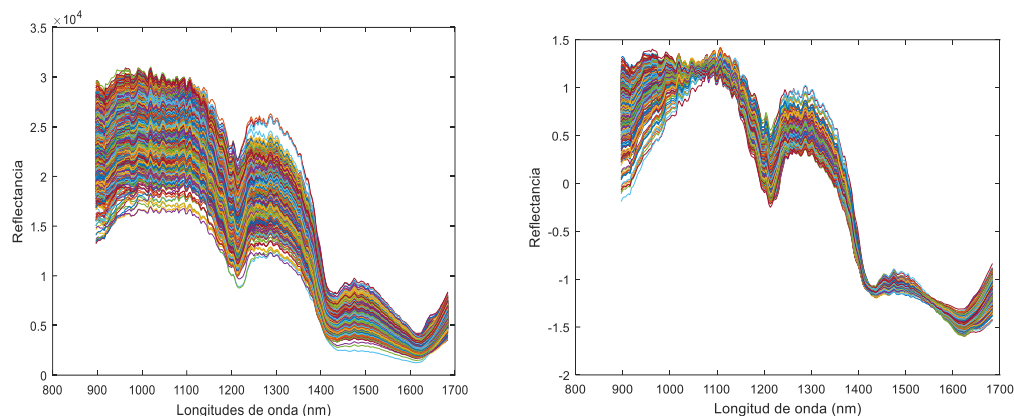


Figura 18 Espectros de aceitunas deshidratadas de la variedad Picual sin pretratamiento, Derecha: Ídem, normalizados.

### 7.1.1 Evaluación por categorías Primera (P), Segunda (S) y Tercera (T).

En este apartado se presentan los resultados de la aplicación del PCA a los espectros de la variedad Picual.

Se puede observar que los loading del PC4 y PC5 son los que muestran pronunciamientos más notorios en las zonas de los 1200 nm y alrededor de los 1400 nm, siendo los otros loadings descartados por presentar demasiado ruido y patrones no muy definidos (Figura 19). Análogamente se muestran los loadings obtenidos del PCA aplicado sobre los espectros normalizados (Figura 20).

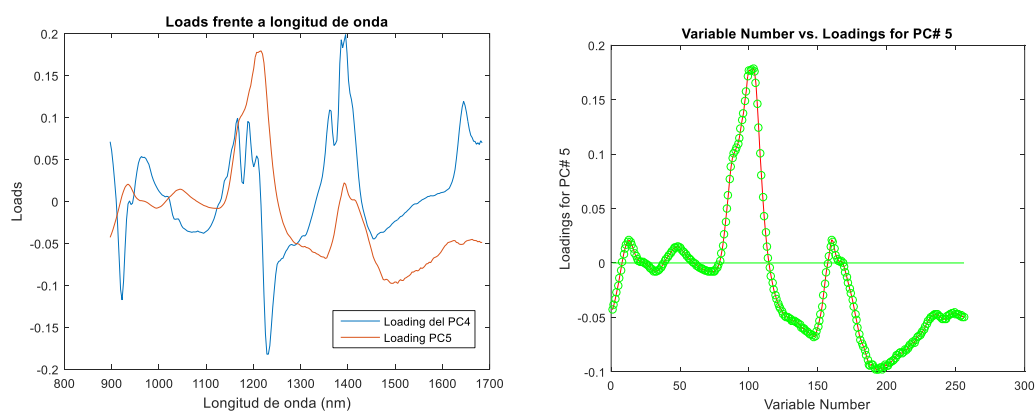


Figura 19 Loadings para espectros sin pretratamiento con valores elevados en torno a 1.200 y 1.400 nm – Variedad Picual.



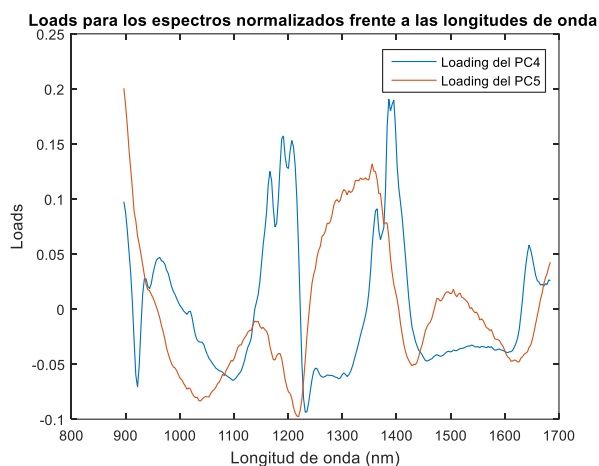


Figura 20 Loading para PCA con espectros normalizados – Variedad Picual.

Al representar los scores sobre el plano definido por PC4 y PC5 vemos que, a pesar de que hay cierta tendencia a segregar según categoría de madurez (P, S y T), no se logra separar totalmente las muestras según estas categorías y existe una gran dispersión. (Figura 21)

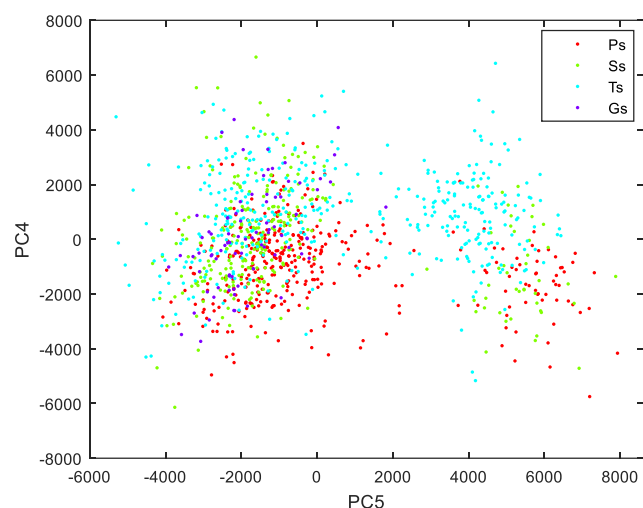


Figura 21 Scores de los PC4 y PC5 de los espectros sin pretratamiento – Variedad Picual.

El score del PC5 sobre las muestras crudas muestra un ANOVA con  $F=79,50$ , que se muestra en la Tabla 6, señalando que existen diferencias significativas para las tres categorías que se están evaluando. La varianza capturada para el modelo de PCA5 es de 0,32 % y una varianza acumulada de 99.84%, y para el PC de los espectros normalizados 0.72% y la varianza acumulada igual a 99.39 % Por otro lado el ANOVA para las muestras normalizadas presenta  $F= 119,74$  evaluado con el scores sobre el PC5. Incrementándose así la el grado de significancia entre grupos. Véase las siguientes tablas.

Tabla 6 (Izquierda) Análisis de varianza para el PC5 de los espectros crudos para determinar la significancia entre categorías; (Derecha) análisis de varianza para el PC5 de espectros normalizados para determinar la significancia entre categorías – Variedad Picual.

ANOVA Table						ANOVA Table					
Source	SS	df	MS	F	Prob>F	Source	SS	df	MS	F	Prob>F
Groups	6.80485e+08	3	2.26828e+08	79.52	1.19202e-46	Groups	5.3367	3	1.77891	119.74	2.76854e-67
Error	3.20323e+09	1123	2.85238e+06			Error	16.684	1123	0.01486		
Total	3.88371e+09	1126				Total	22.0207	1126			

La Figura 22 presenta la media y la dispersión de las muestras para las tres categorías, siendo la categoría “primera” (P) la que presenta menor variabilidad frente a las categorías de segunda (S) y Tercera (T). Y para las muestras normalizadas en la Figura 22 (derecha) se distingue una reducción de la dispersión entre las muestras respecto a su media, siendo la segunda categoría la que presenta mayor dispersión dentro del grupo.

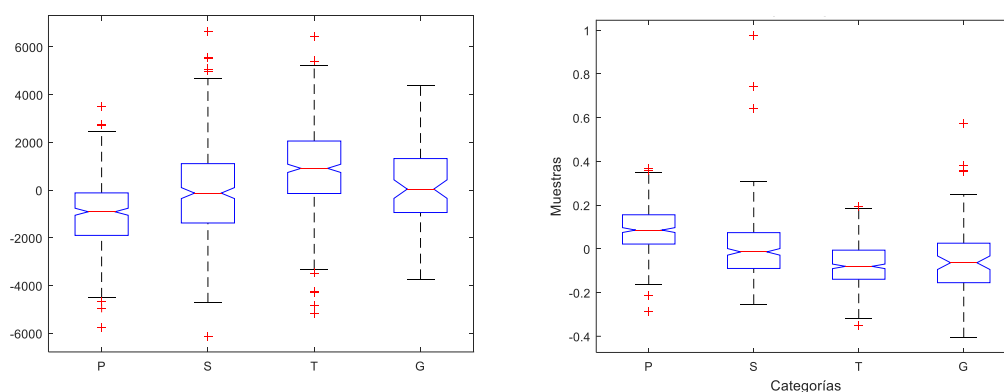


Figura 22 Gráfico de Cajas; (Izquierda) para el score 5 sobre las muestras crudas; (Derecha) Ídem sobre los datos normalizados – Variedad Picual.

La Figura 23 muestra la comparación múltiple de las categorías P, S y T para los scores del PC5 sobre los espectros crudos y normalizados respectivamente. La media de la categoría P se pone de relieve y el intervalo de comparación, de color azul. Debido a que los intervalos de comparación para los otros dos grupos no se cruzan con los intervalos para la media del grupo P, que se destacan en rojo, indica que existen diferencias significativas para las tres categorías.

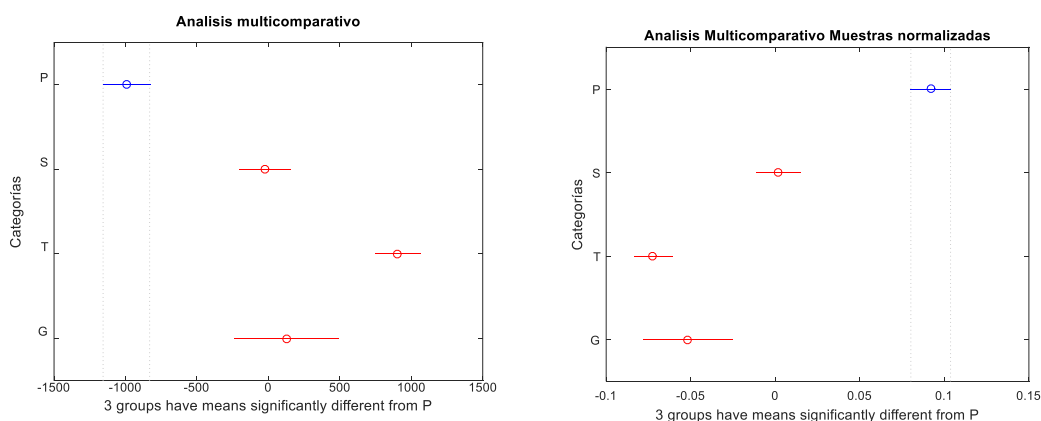


Figura 23 (Izquierda) Comparación múltiple de los datos estadísticos del ANOVA sobre el PC5 de los espectros crudos para determinar si existen diferencias significativas entre las categorías P, S y T; (Derecha) Ídem sobre los espectros normalizados – Variedad Picual.

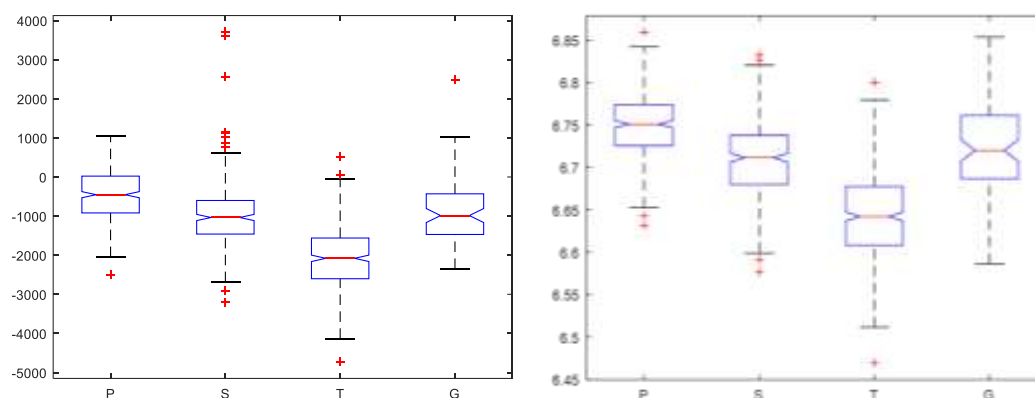
A continuación se expone el resultado de la aplicación del análisis PLSDA. Tras realizar el análisis PLSDA y realizar el ANOVA sobre los scores obtenidos con el vector de proyección que considera 6 variables latentes, se obtiene la Tabla 7, que muestra que existe una F alta,  $F=311,74$ , lo que quiere decir que existen diferencias significativas entre los grupos P, S y T. Y al ser un análisis dirigido muestra mejores resultados de segregación que el análisis PCA. Mientras que para los datos normalizados se logra una mayor diferencia significativa, F igual a 388.

*Tabla 7 Izquierda - ANOVA sobre los scores del PLSDA en los espectros crudos; Derecha - Ídem sobre los espectros normalizados – Variedad Picual.*

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	5.47243e+08	3	182414411.5	311.74	3.30361e-147
Error	6.57121e+08	1123	585148.2		
Total	1.20436e+09	1126			

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	2.31682	3	0.77227	388.96	3.42647e-173
Error	2.2297	1123	0.00199		
Total	4.54653	1126			

La Figura 24 Izquierda muestra claramente cómo con el análisis de PLSDA para los espectros crudos se minimiza la dispersión de las muestras respecto a su media para las 3 categorías y permiten segregar las categorías P, S y T. Y para los espectros normalizados se refleja una segregación más notoria entre las categorías, minimizando la dispersión dentro de la población entre categorías.



*Figura 24 Izquierda: Gráfico de cajas para el análisis PLSDA con espectros crudos; Derecha: Gráfico de cajas para el análisis PLSDA sobre los espectros normalizados – Variedad Picual.*

Por otro lado la Figura 25 muestra que la comparación múltiple presenta una gran separación entre los tres grupos, apreciándose que existe mayor distancia entre los grupos S y T tanto para las muestra crudas (Izquierda), como para las normalizadas (Derecha).

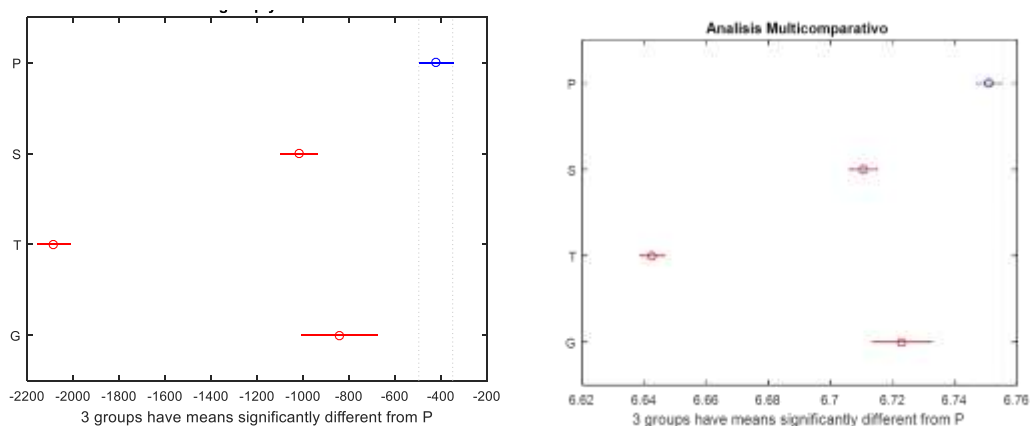


Figura 25 Izquierda: Análisis de Comparación múltiple para los scores obtenidos mediante PLSDA sobre los espectros crudos para las categorías P, S y T; Derecha. Ídem para espectros normalizados. – Variedad Picual.

### 7.1.2 Evaluación por categorías: Buenas (B) y Dañadas (D).

Para comprobar si los espectros NIR son sensibles al estado de calidad de los frutos (con o sin defectos), se realizó un PCA sobre el conjunto de la población de los frutos frescos, considerando la categorización de B (buenas) y D (con defectos). Como muestra la Figura 26, el PC5 (la varianza capturada para el modelo de PCA 5 es de 0,13%. y la varianza acumulada de 99,83%,) presenta valles más pronunciados en las regiones de 1200 y 1400nm, relacionado entonces el estado del fruto con el contenido de aceite y de agua. Es el PC5, el que muestra mayores diferencias entre las categorías B y D, según los ANOVA realizados sobre los scores obtenidos de este PCA, con una F de 361 como muestra la Tabla 8.

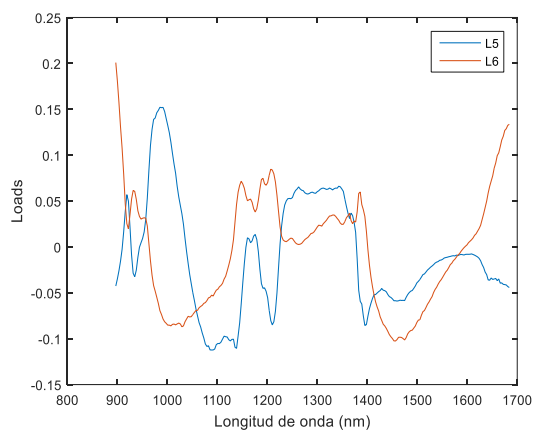


Figura 26 Loading correspondiente a la aplicación de ACP sobre los espectros no tratados de aceitunas frescas – Variedad Picual.

Tabla 8 cuadro del ANOVA Realizado sobre los scores del PC #5 – Variedad Picual

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	3.50613e+08	1	350612888	361.34	1.38333e-71
Error	1.3109e+09	1351	970314.6		
Total	1.66151e+09	1352			

En la Figura 27 se incluye el gráfico de cajas, donde se muestran las medias y la dispersión de las poblaciones de las dos categorías, la categoría B presenta menor dispersión de las muestras respecto a su media, siendo la categoría D la de mayor dispersión de muestras respecto a su media.

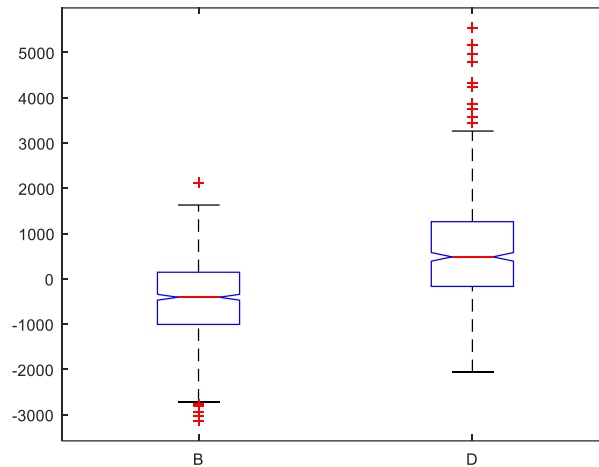


Figura 27 Gráfico de cajas del ANOVA Realizado sobre los scores del PC5 para determinar la significancia entre las categorías B y D – Variedad Picual.

El análisis PLSDA sobre los scores obtenidos del PCA muestra una alta significancia para estas mismas categorías siendo  $F= 1353,94$ . (Tabla 9).

La Figura 28 presenta el vector de proyección para el análisis PLSDA, donde se aprecian valles más pronunciados e influyentes para la segregación entre los grupos B y D. Resalta el PC5 para la variedad Picual.

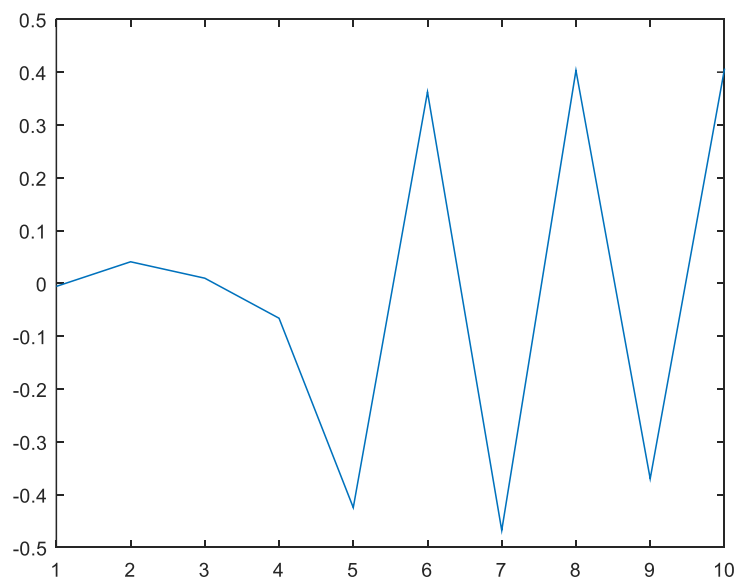


Figura 28 Vector de proyección para el análisis PLSDA – Variedad Picual.

Tabla 9 Cuadro del ANOVA realizado sobre los scores del PLSDA – Variedad Picual.

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	3.68318e+08	1	368317695.8	1353.94	6.6383e-206
Error	3.67519e+08	1351	272034.7		
Total	7.35837e+08	1352			

El gráfico de cajas (Figura 29) indica las diferencias entre las medias y las dispersiones de las poblaciones de cada categoría. En la categoría B se presentan las muestras dentro de la caja ligeramente dispersas respecto a su media, sin embargo también se puede observar que existen muestras que escapan del grupo tomado dentro de la caja por ser valores muy extremos. Para la categoría D se presentan las muestras ligeramente más dispersas dentro de la caja y menor número de muestras de valores extremos (signos en forma de cruz roja).

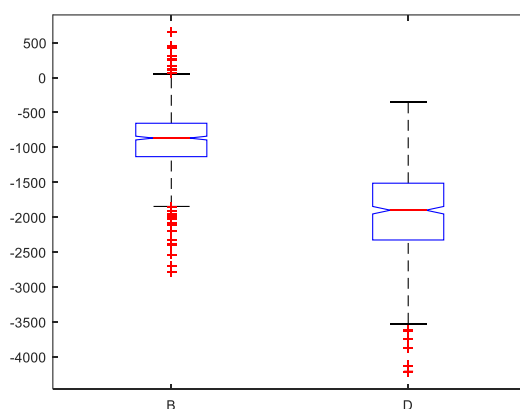


Figura 29 Gráfico de cajas para el ANOVA realizado sobre los scores del PLSDA, para determinar la significancia entre B y D – Variedad Picual

La proyección de los espectros (tanto de la muestra con y sin daño) en el espacio de los componentes principales PC5 y PC7, permite diferenciar los grupos solapados, esto debido a la gran dispersión (Figura 30).

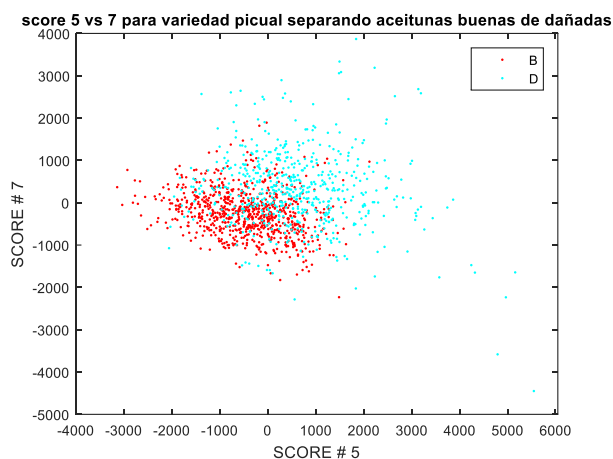


Figura 30 Scores de los PC5 y PC7 de los espectros sin pretratamiento de las categorías B y D – Variedad Picual

### 7.1.3 Rendimiento graso

En la Ilustración 22 Izquierda se incluye el gráfico de cajas del rendimiento graso según el día de recolección. Se observa una evolución hacia valores crecientes del contenido graso según avanza la campaña. Tratando de buscar una relación entre la información contenida en los espectros NIR y el contenido graso de las muestras (valores medios), sobre los scores del PC5 obtenidos de los espectros sin normalizar, se aplicó un ANOVA considerando como factor el día de recolección (Ilustración 22, Derecha). Se observa que los valores medios de los scores evaluados presentan la misma tendencia o comportamiento que los valores del contenido graso, pero en sentido inverso (los scores decrecen según aumenta el contenido graso). El día 1 presenta el menor valor de la media y a medida que trascurren los días de recolección va incrementándose el valor de la media, teniendo el día 9 como el de mayor valor de rendimiento graso. Esta similitud de comportamiento permitiría relacionar la información aportada por los espectros NIR con el contenido graso en las aceitunas. Hay que indicar que esta relación a lo largo de las fechas de recolección, incluye también la relación con las categorías de madurez, porque según avanza la campaña las categorías que prevalecen son las categorías S y T, desapareciendo la categoría P después de las dos primeras fechas de recolección. (Ver Figura 31).

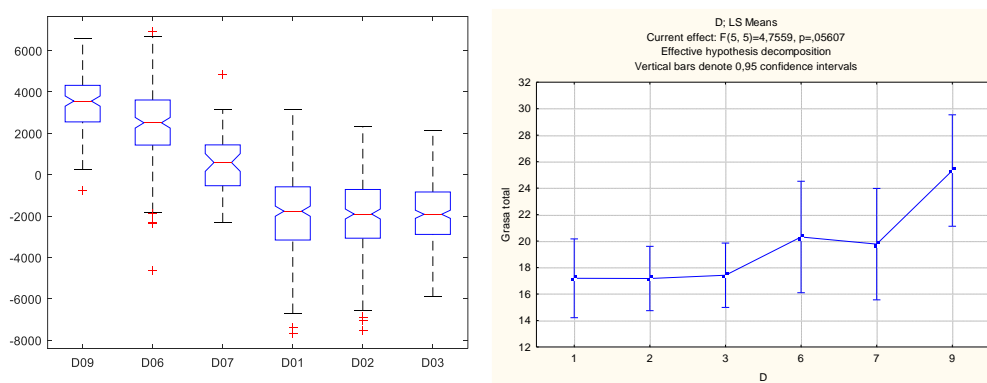


Figura 31 (Izquierda) ANOVA del PC5 para determinar la significancia entre los días de recolección de las aceitunas de almazara; (Derecha) Representación de los días de recolección de aceitunas de almazara frente al contenido de grasa total del análisis de referencia – Variedad Picual

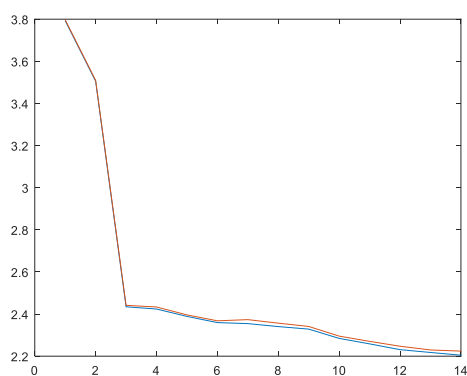
En la Tabla 10 se observa que existen diferencias significativas, representado por  $F=330$  y el valor de  $P$  cercano a cero, entre los scores del PC5 según el factor día de recolección de aceituna de almazara.

*Tabla 10 Análisis de varianza del PC5 en función a los días de recolección para los espectros crudos de la variedad Picual*

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	4.97854e+09	5	9.95707e+08	330.25	1.82782e-217
Error	3.37986e+09	1121	3.01504e+06		
Total	8.3584e+09	1126			

#### 7.1.4 Contenido de humedad

Para determinar el contenido de humedad se realizó una la regresión multilíneal para los frutos de Picual; las primeras 7 de las 14 longitudes de onda incluidas en el modelo de estimación del contenido en humedad se incluyen en la Tabla 11; se muestran estas 7 longitudes de onda a partir del gráfico de los errores estándar de calibración y de validación, frente al número de variables incluidas en el modelo (Figura 32).



*Figura 32 Representación de los errores de calibración (SEC) y error de validación cruzada (SECV) frente al número de longitudes de onda incluidas en el modelo – Variedad Picual*

*Tabla 11 Longitudes de onda seleccionadas en la regresión multilíneal paso a paso para la estimación del contenido en humedad – Variedad Picual*

Longitud de onda (nm)	1398	1238	1200	1194	1184	1560	1138
-----------------------	------	------	------	------	------	------	------

La regresión multilíneal alcanza un coeficiente de determinación igual a  $R^2 = 0,63$  como muestra la Figura 33 (Izquierda) tomando las tres primeras longitudes de onda seleccionadas en el proceso paso a paso de la correlación entre los errores de calibración (SEC) y error de validación cruzada (SECV). Esto explica que las longitudes de onda relacionadas a la humedad están alrededor de 1200 y 1398nm. Por otro lado al tomar las 7 longitudes de onda no se logra mejorar mucho el coeficiente de determinación ( $R^2 = 0,65$ ) como muestra la Figura 33 (Derecha)



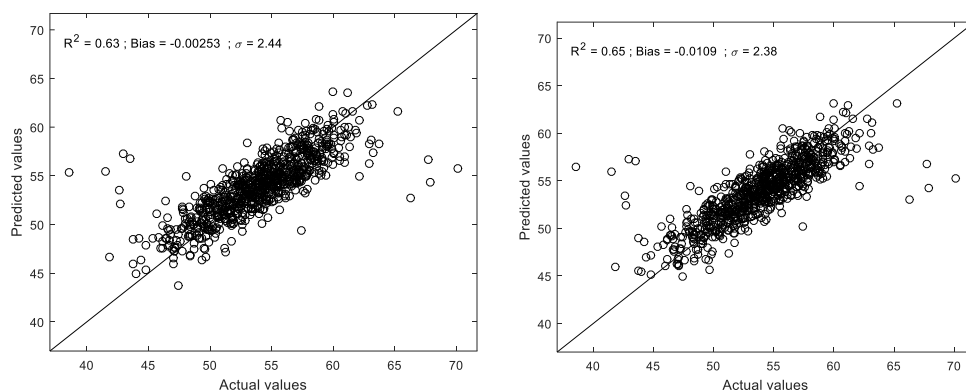


Figura 33 Izquierda) Modelo de Regresión lineal múltiple del contenido de humedad para la variedad Picual tomando 3 Longitudes de onda, (Derecha) Ídem tomado 7 longitudes de onda.

## 7.2 Variedad Cornicabra

La Figura 34 muestra el conjunto de los espectros de la variedad Cornicabra ( $n=5894$ ); a la derecha se presentan los espectros normalizados y se puede observar que existe gran dispersión, a la izquierda se puede apreciar los espectros normalizados con menor dispersión y valles pronunciados alrededor de los 1200 nm y los 1400 nm.

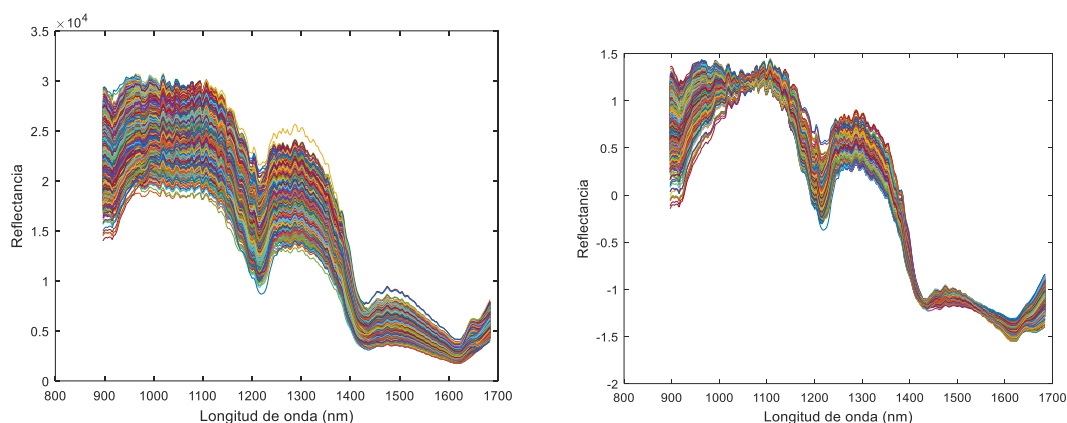


Figura 34 Izquierda: Espectros de aceitunas deshidratadas de la variedad Cornicabra sin pretratamiento, Derecha: Ídem, normalizados

### 7.2.1 Evaluación por categorías Primera (P), Segunda (S) y tercera (T).

Un procedimiento de análisis análogo al mostrado en la variedad Picual se siguió en la variedad Cornicabra. En primer lugar se realizó sendos PCA sobre los espectros crudos (sin pretratamiento) y los espectros normalizados. En la Figura 35 (Izquierda) se representa los loadings del PC5 Y PC6, los dos presentan valles pronunciados en la región de los 1200nm, sin embargo el PC5 muestra ruido, por tal motivo se usa el PC6 (Derecha) se muestra que los loading del PC5 y PC6 para los espectros normalizados, sin embargo el PC5 al igual que los demás espectros presenta mucho ruido por lo que se opta por el PC6.

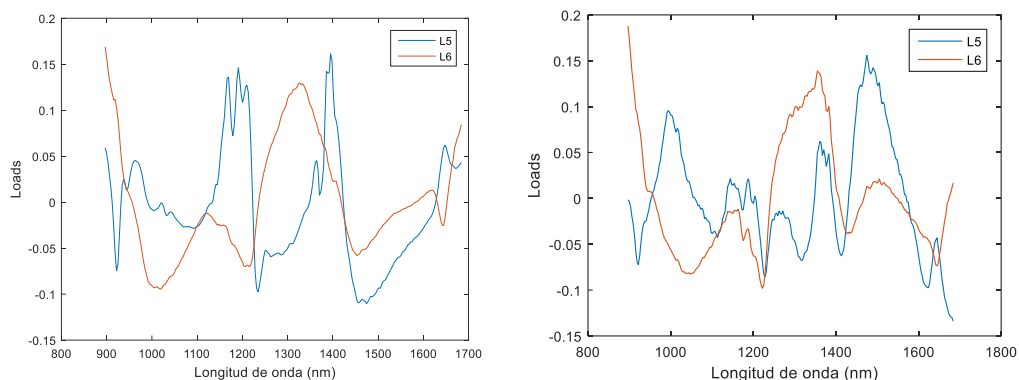


Figura 35 (Izquierda) Loadings para espectros sin pretratamiento con valores elevados en torno a 1.200 y 1.400 nm; (Derecha) Ídem normalizados. - Var Cornicabra.

El componente principal más relacionado con las categorías de madurez para los espectros sin normalizar y normalizados es el PC6. Los resultados de los análisis de varianza (ANOVA), sobre los scores en estos PC's, realizados para comprobar si existen diferencias significativas entre las categorías P, S y T se muestran en la Tabla 12: existen diferencias significativas, tanto para los espectros no normalizados como para los normalizados, con valores de F de 195, 29 y 180 respectivamente.

Siendo la varianza capturada para el modelo de PCA 6 0.18% y la varianza acumulada igual a 99.87%, mientras que para los espectros normalizados es de 0.39%. Y la varianza acumulada de 99.44%.

Tabla 12 (Izquierda) Cuadro del ANOVA sobre el PC 6 con espectros no normalizados, con el factor categoría de madurez; (Derecha) Ídem sobre espectros normalizados – Variedad Cornicabra.

ANOVA Table						ANOVA Table					
Source	SS	df	MS	F	Prob>F	Source	SS	df	MS	F	Prob>F
Groups	6.2676e+08	3	2.0892e+08	195.29	3.62633e-115	Groups	7.4936	3	2.49787	180.59	3.44664e-107
Error	3.04574e+09	2847	1.06981e+06			Error	39.3788	2847	0.01383		
Total	3.6725e+09	2850				Total	46.8724	2850			

El gráfico de cajas (Figura 36, izquierda) muestra que las medias de las tres categorías se diferencian claramente, además presenta baja dispersión de muestras, siendo las categorías S y T las que poseen más muestras en valores extremos (marcas en rojo). El gráfico de cajas (Figura 36 derecha) muestra las medias de las categorías evaluadas y una mayor dispersión de las muestras dentro de cada categoría, lo que explica claramente por qué el valor de F para las muestras normalizadas es menor que el de las muestras sin normalizar. Otro factor observado en la Figura 36 (Derecha) es la dispersión de la categoría P respecto a su media.

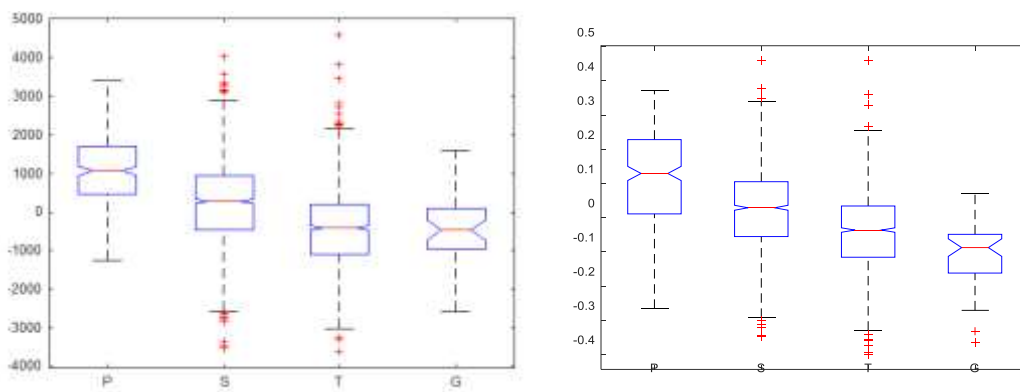


Figura 36 Gráfico de Cajas; (Izquierda) para el score 6 sobre las muestras crudas; (Derecha) Ídem sobre los espectros normalizados - Variedad Cornicabra.

Posteriormente se realizó un análisis de comparación múltiple, con los datos estadísticos obtenidos del ANOVA, para determinar qué categorías son significativamente distintas entre sí. La Figura 37 (Izquierda) muestra que existen diferencias significativas entre todos los grupos P, S y T. La imagen derecha muestra que también existen diferencias significativas entre las tres categorías para las muestras normalizadas.

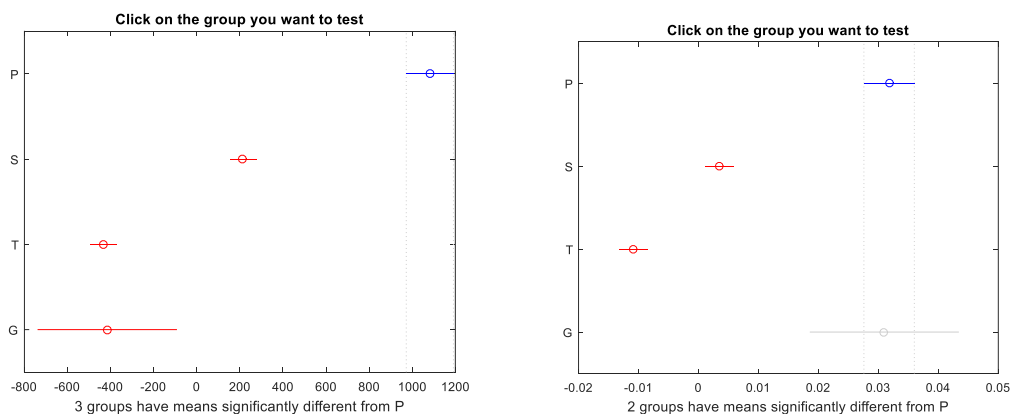


Figura 37 (Izquierda) Comparación múltiple de los datos estadísticos del ANOVA sobre el PC6 de los espectros crudos para determinar si existen diferencias significativas entre las categorías P, S y T; (Derecha) Ídem sobre los espectros normalizados – Variedad Cornicabra.

Otro análisis que también se realizó para determinar las diferencias significativas entre las categorías P, S y T fue el PLSDA, considerando los 10 primeros scores del análisis de componentes principales, con su respectivo ANOVA y análisis de comparación múltiple.

La Tabla 13 (Izquierda) muestra que existen diferencias significativas entre las categorías evaluadas ( $F=569,8$ ), mejorando la significancia presentada por el PC6 ( $F=195,29$ ) a través del análisis PLSDA la segregación entre categorías. Para las muestras normalizadas (Derecha) se obtiene un valor de  $F=541,41$ , indicando que existen diferencias significativas entre las categorías.

Tabla 13 (Izquierda) ANOVA sobre los scores del PLSDA en los espectros crudos; (Derecha) Ídem sobre los espectros normalizados – Variedad Cornicabra.

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	4.01071e+08	3	133690491.5	569.8	4.87227e-290
Error	6.6798e+08	2847	234625.8		
Total	1.06905e+09	2850			

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	4.0403	3	1.34678	541.43	2.19177e-278
Error	7.0818	2847	0.00249		
Total	11.1222	2850			

El gráfico de cajas (Figura 38 Izquierda) muestra que las medias de las categorías P, S y T están claramente diferenciadas entre grupos, exponiendo un solape de muestras entre las categorías S y T. El gráfico de la derecha muestra las categorías con sus respectivas medias para las muestra normalizadas, donde se puede apreciar que para la segunda categoría las muestras están más dispersas respecto a la media, lo que explicaría un menor valor de F respecto a la del ANOVA para las muestras sin normalizar.

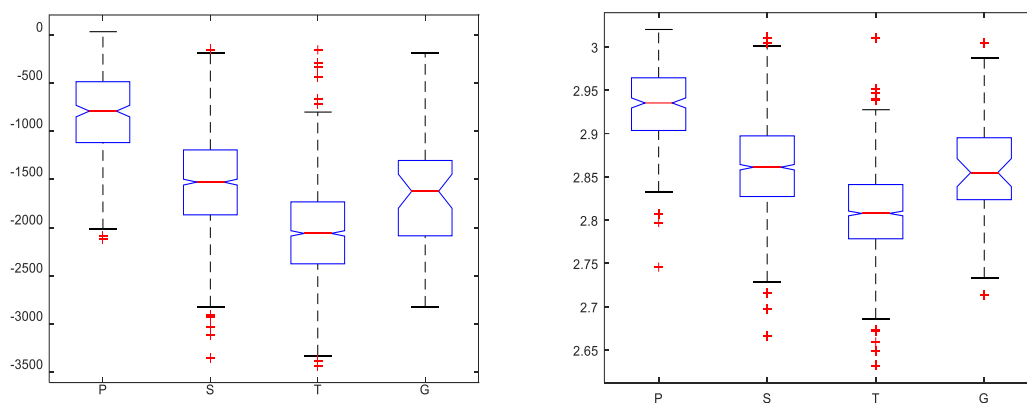


Figura 38 (Izquierda) Gráfico de cajas para el análisis PLSDA con espectros crudos; (Derecha) Ídem sobresobre los espectros normalizados - Variedad Cornicabra.

La Figura 39 (Izquierda) muestra la proyección de la totalidad de la población en el plano de los dos vectores discriminantes del PLSDA, se observa que es el a lo largo del primer vector discriminante donde se produce una ordenación de los individuos según categoría de madurez, aunque existe una gran dispersión entre las muestras de una misma categoría, La Figura 39 de la derecha muestra que después de normalizar los datos no se logra una mejor segregación de las categorías.

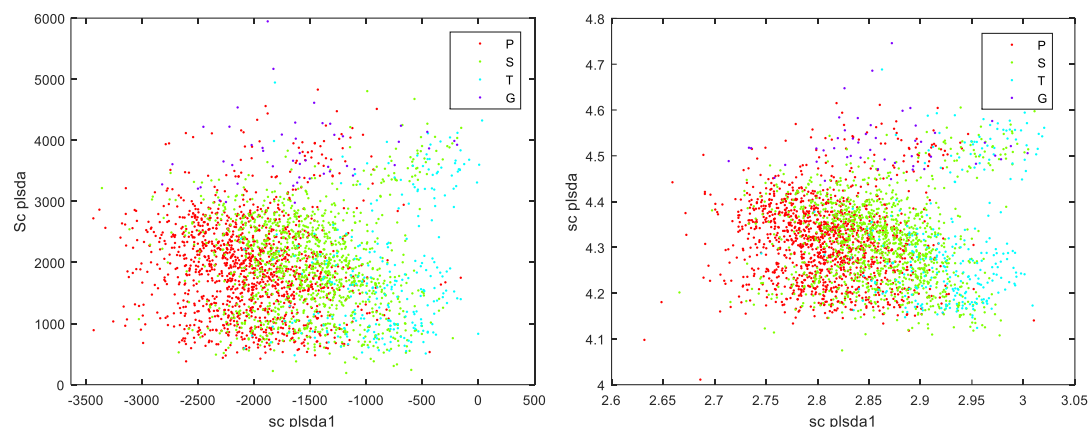


Figura 39 (Izquierda) Scores 1 y 2 del PLSDA de los espectros sin pretratamiento, (Derecha) Ídem en espectros normalizados – Variedad Cornicabra.

### 7.2.2 Evaluación entre categorías de frutos buenos (B) y dañados (D)

A continuación se presenta los resultados obtenidos para la aplicación del PCA sobre los espectros sin pretatamiento, tomados sobre aceitunas frescas ( $n = 3043$ ), de la variedad Cornicabra.

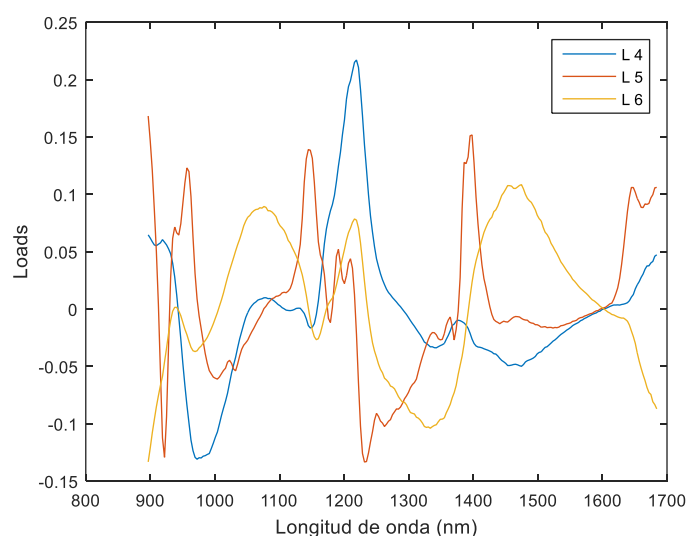


Figura 40 Loadings para espectros sin pretratamiento con valores elevados en torno a 1.200 y 1.400 nm – Variedad Cornicabra.

La Figura 40 muestra los PC4, PC5 y PC6 que tienen valles pronunciados alrededor de los 1200 y 1400; tanto el PC5 y PC6 muestran mucho ruido, por tal motivo se opta por el PC4 que presenta una varianza capturada de 0,31%. Y la varianza acumulada de 99,39%,

El análisis de PCA para determinar las categorías indica que el PC4 es el más representativo y contiene mejor información de los espectros, como muestra la Figura 40 al desarrollar el ANOVA obtenemos la Tabla 14 que muestra el valor de  $F = 645,49$  representando que existe alta significancia entre las categorías B y D. Con una probabilidad casi nula de que no se vuelva a presentar el valor de  $F$  tan alto.

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	1.45907e+09	1	1.45907e+09	645.49	2.68822e-129
Error	6.87391e+09	3041	2.26041e+06		
Total	8.33298e+09	3042			

Tabla 14 Cuadro del ANOVA Realizado sobre los scores del PC 4 – Variedad Cornicabra.

El gráfico de cajas que se presenta en la Figura 41 muestra que para la categoría D existen elementos fuera del rango de consideración de los valores pertenecientes a las cajas (representados por signos en forma de cruz en color rojo), por ser muy extremos respecto a los valores medios.

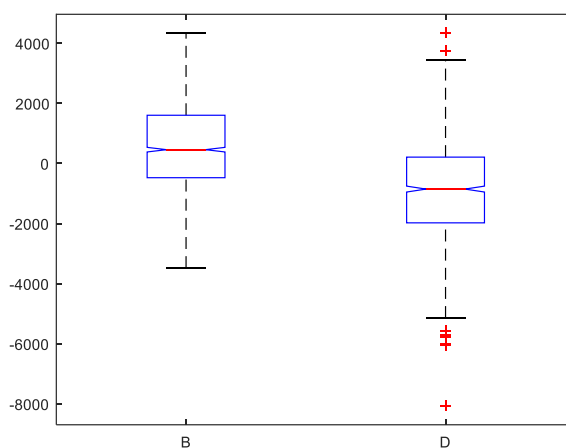


Figura 41 Gráfico de cajas del ANOVA Realizado sobre los scores del PC4 para determinar la significancia entre las categorías B y D – Variedad Cornicabra.

Tras el PCA se realizó el PLSDA sobre los scores de los 10 primeros PC, se puede apreciar que el vector de proyección indica que además del PC4 el PC8 tiene un peso importante en la segregación entre B y D. El correspondiente análisis de varianza, resultó en una  $F=1556,82$  con diferencias altamente significativas entre las categorías B y D. El valor de F se ha incrementado notablemente con el análisis PLSDA. (Tabla 15)

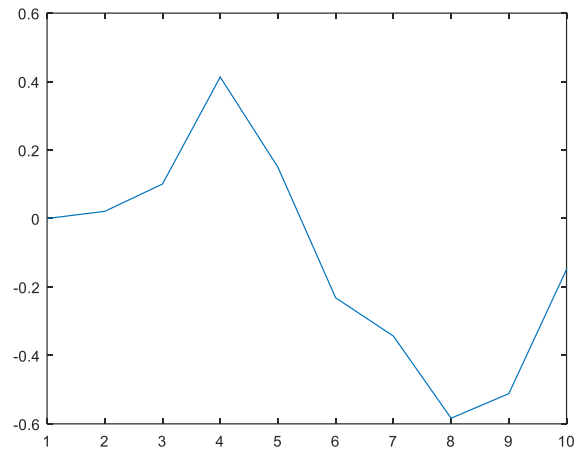


Figura 42 Vector de proyección para el PLSDA – Variedad Cornicabra.

Tabla 15 ANOVA realizado sobre los scores del PLSDA considerando como factor la presencia de defectos – Variedad Cornicabra.

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	9.2753e+08	1	927530364.9	1556.82	2.57675e-275
Error	1.81178e+09	3041	595784.1		
Total	2.73931e+09	3042			

El gráfico de cajas muestra que las medias entre las categorías son claramente diferentes entre sí y muestran una menor dispersión de las muestras respecto a su media. También se puede observar que el grupo B presenta mayor número de muestras de valores extremos ubicados en la parte externa - inferior de la caja, esto permite que dentro del grupo B se presente menor dispersión al no considerar estos valores (Figura 43).

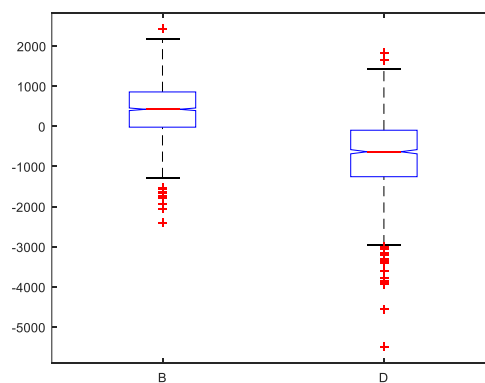


Figura 43 Gráfico de cajas para el ANOVA realizado sobre los scores del PLSDA, para determinar la significancia entre B y D – Variedad Cornicabra.

### 7.2.3 Rendimiento graso

Para el contenido de aceite se realizó un ANOVA sobre este contenido graso según los análisis de referencia, considerando la fecha de recolección como factor. Este ANOVA indica que existen diferencias significativas entre los días de recolección ( $F=33,8$ ;  $p<0,05$ ). Para comprobar si ese mismo patrón se observa en el PC más relacionado con categoría de madurez se realizó un ANOVA sobre el mismo considerando como factor la fecha de recolección; aunque se observan diferencias significativas ( $F=913$ , Tabla 16), no se replica el patrón de comportamiento que se observa en el contenido graso a lo largo de la campaña, como sí sucedía en el caso de Picual. (Ver Figura 44).

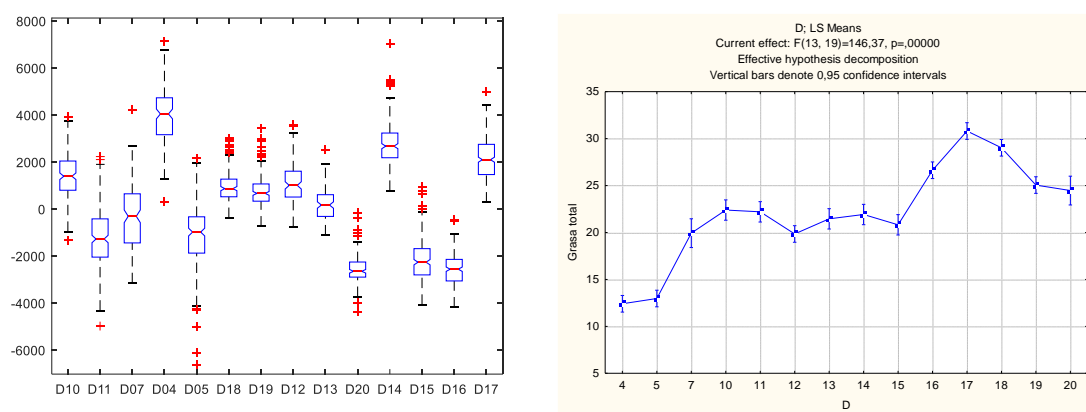


Figura 44 (Izquierda) Gráfico de cajas para el ANOVA con el PC6 y su incidencia en los días de recolección y (Derecha) Representación de la variación del contenido de grasa en los días de recolección – Variedad Cornicabra.

Así, en el caso de Cornicabra el patrón de las medias de los scores representado para los días de recolección no coincide con el patrón del contenido de grasa total representado para las muestras de referencia. Esta no coincidencia puede deberse a las zonas de procedencia que fue muy variada, como se indica en la Tabla 3 de la procedencia de muestras de aceituna.

Tabla 16 Cuadro de ANOVA para la incidencia de los días de recolección – Variedad Cornicabra.

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	1.10841e+10	13	852622350.8	913.27	0
Error	2.6486e+09	2837	933591.2		
Total	1.37327e+10	2850			

En la Tabla 16 se observa que existen diferencias significativas, representado por  $F=913$  y el valor de  $p=$  cero, entre los scores del PC5 según el factor día de recolección de aceituna de almazara.

### 7.2.4 Contenido de humedad

Para determinar el contenido de humedad se realizó una la regresión multineal para los frutos de Cornicabra; las primeras 10 de las 14 longitudes de onda incluidas en el modelo de estimación del contenido en humedad se incluyen en la Tabla14; se muestran estas 10 longitudes de onda a



partir del gráfico de los errores estándar de calibración y de validación, frente al número de variables incluidas en el modelo (Figura 45).

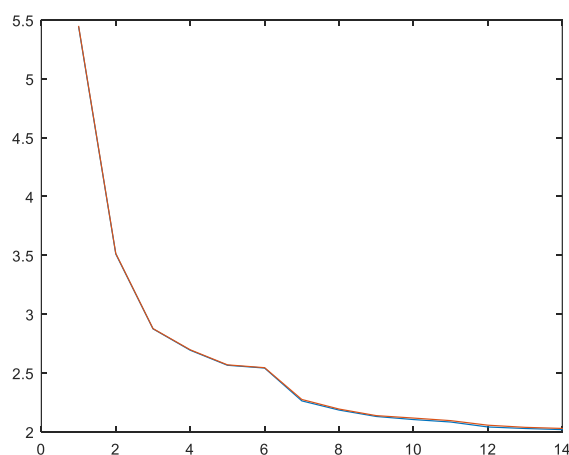


Figura 45 Representación de los errores de calibración (SEC) y error de validación cruzada (SECV) frente al número de longitudes de onda incluidas en el modelo – Variedad Cornicabra.

Para cornicabra la regresión multilineal presenta valores de predicción altos, obteniendo una  $R^2 = 0,84$ , acercando el modelo a una buena predicción para el contenido de humedad.

Tabla 17 Longitudes de onda seleccionadas en la regresión multilineal paso a paso para la estimación del contenido en humedad – Variedad Cornicabra.

Wave (nm)	1352	1209	988	1629	1197	1054	1321	1416	1178	1022
-----------	------	------	-----	------	------	------	------	------	------	------

La regresión multilineal alcanza un coeficiente de determinación igual a  $R^2 = 0,84$  como muestra la Figura 46, tomando las siete primeras longitudes de onda seleccionadas en el proceso paso a paso de la correlación entre los errores de calibración (SEC) y error de validación cruzada (SECV). Esto explica que las longitudes de onda relacionadas a la humedad están alrededor de 988 y los 1629 nm.

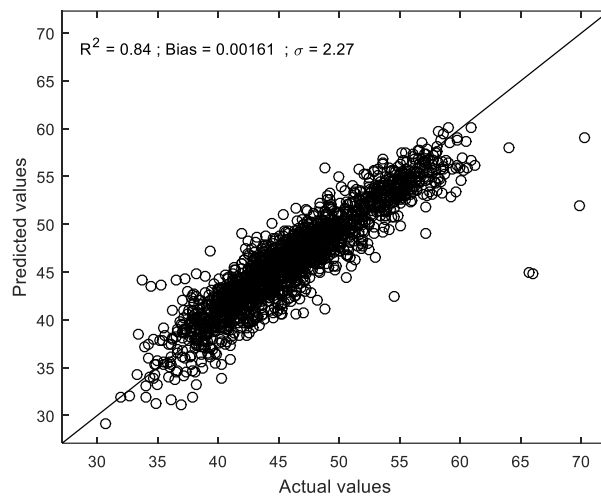


Figura 46 Modelo de Regresión lineal múltiple del contenido de humedad para la variedad Cornicabra.

Vemos que existe un buen ajuste entre los valores actuales y la predicción para este modelo.

El modelo predictivo para el contenido de humedad ha dado resultados aceptables para las variedades Cornicabra y Picual, obteniéndose valores de  $R^2$  similares a los obtenidos por Gracia y León (2011), recomendándose la validación del modelo con datos obtenidos sobre aceitunas de distintas procedencias.

## 8 CONCLUSIONES

En el presente trabajo se ha aplicado la espectroscopia NIR para la evaluación de aceitunas intactas recogidas en la recepción de una almazara.

Se han podido establecer modelos con coeficientes de determinación superiores a 0,6 para la variedad Picual y superiores a 0,8 en Cornicabra, para la estimación del contenido en humedad de las aceitunas, sobre la base de que la referencia para esta variable se toma aceituna a aceituna, esto es, de cada fruto del que se toma el espectro NIR se conoce cuál es su porcentaje de humedad en el momento en el que se toma dicho espectro. Ofreciendo la posibilidad de implementar esta técnica en las industrias oleicas antes del molturado, y de esta manera obtener aceites de mejor calidad, partiendo de la relación agua – contenido de grasa. Y disminuir los procesos de separación de la fracción oleaginosa del agua.

En cuanto a la estimación del contenido graso de las aceitunas analizadas, la mayor dificultad se ha encontrado en el hecho de que como referencia de esta variable, sólo se tiene el valor agrupado correspondiente a un conjunto de más de 200 g de aceitunas tomadas del mismo lote que de la que se toman las aceitunas sobre las que se determina el espectro NIR. Por ello en este trabajo se ha optado por considerar la categorización de los frutos según su nivel de madurez en lugar de estimar el contenido graso, y evaluar la capacidad de la espectroscopia NIR para segregar entre 3 niveles de madurez (verdes y pintonas, moradas y negras) y 2 niveles de calidad (sin y con defectos). Aun así se puede concluir que la técnica NIR permite desarrollar un buen modelo de segregación en base a los análisis de referencia que indican que la segunda categoría (S) es la que obtiene mayor porcentaje graso.

Se ha comprobado que el análisis de componentes principales permite la identificación de los niveles de madurez referidos mediante componentes cuyos loadings tienen un peso muy marcado alrededor de 1.200 nm, longitud de onda relacionada con las grasas, al corresponder con el segundo sobretono de la vibración de tensión del enlace C-H. Las diferencias entre niveles de madurez se acentúan, para las dos variedades estudiadas, al aplicar un PLS-DA sobre los scores generados con el análisis de componentes principales.

Resultados similares se observan al tratar de segregar las aceitunas según su nivel de calidad.

Con el objetivo de tratar de mejorar la capacidad de la espectroscopia NIR para la evaluación de la composición de frutos pequeños como las aceitunas, se hace necesario la adecuación de los métodos de referencia para obtener valores referidos a cada fruto.

## 9 REFERENCIAS

- Abbott, J. A. (1999). Quality measurement of fruits and vegetables. *Postharvest Biology and Technology*, 15(3), 207-225.
- Alcalà Bernàrdez, M. (2006). Utilización de la espectroscopia NIR en el control analítico de la industria farmacéutica. desarrollos iniciales en PAT.
- Aparicio, R., Morales, M. T., Aparicio-Ruiz, R., Tena, N., & García-González, D. L. (2013). Authenticity of olive oil: Mapping and comparing official methods and promising alternatives. *Food Research International*, 54(2), 2025-2038.
- Armenta, S., Moros, J., Garrigues, S., & Guardia, M. D. L. (2010). The use of near-infrared spectrometry in the olive oil industry. *Critical Reviews in Food Science and Nutrition*, 50(6), 567-582.
- Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1993). *Correction to the description of standard normal variate (snv) and de-trend transformations in practical spectroscopy with application in food and beverage analysis* (2nd ed.) Journal of Near Infrared Spectroscopy.
- Bautista Mercader, M., & Blanco Romia, M. (2009). Avances en la aplicación de la espectroscopía NIR en la industria farmacéutica. Introducción a PAT y técnicas de imagen.
- Brereton, R. (2007). In University of Bristol U. (Ed.), *Applied chemometrics for scientists* John Wiley & Sons Ltd Chichester.
- Cayuela JA, Garcia JM, Caliani N. (2009). NIR prediction of fruit moisture, free acidity and oil content in intact olives. *Grasas Y Aceites*, 60(2), 194-202. doi:10.3989/gya.097308
- Delgadillo, I., Nunes, A., & Barros, A. (2012). *Quality evaluation of olives, olive pomace and olive oil by infrared spectroscopy* INTECH Open Access Publisher.
- Essiari, M., Zouhair, R., & Chimi, H. (2014). Contribución al estudio de la tipicidad de los aceites de oliva vírgenes producidos en la región de saïs (marruecos). *Revista Oficial Del Consejo Oleícola Internacional*, 119
- Frankel, E. N. (2010). Chemistry of extra virgin olive oil: Adulteration, oxidative stability, and antioxidants. *Journal of Agricultural and Food Chemistry*, 58(10), 5991-6006.
- Gigahertz-Optik. (2016). *Tutorial - basics of light measurement*
- Gomez, G., González, M., Pérez, R., Pérez, A., Lazaro, A., & De Lorenzo, C. (2004). Alteraciones y pérdida de calidad en aceituna de mesa y aceite de oliva. *Enciclopedia Del Olivo - Centros De Difusión Tecnológica - TDC-Olive*,
- Gorry, P. A. (1990). General least-squares smoothing and differentiation by the convolution (savitzky-golay) method. *Analytical Chemistry*, 62(6), 570-573.
- Gracia, A., & León, L. (2011). Non-destructive assessment of olive fruit ripening by portable near infrared spectroscopy. *Grasas Y Aceites*, 62(3), 268-274.

- Greensill, C. V., & Walsh, K. B. (2000). A remote acceptance probe and illumination configuration for spectral assessment of internal attributes of intact fruit. *Meas. Sci. Technol*, 11
- Grupo de Quimiometría y Cualimetría de Tarragona. (2002). QUIMIOMETRÍA, una disciplina útil para el análisis químico. *Técnicas De Laboratorio*, 24(272), 412-416.
- Guzmán, E., Baeten, V., Fernández, J. A., & García-Mesa, J. A. (2013). Infrared machine vision system for the automatic detection of olive fruit quality. *Talanta*, 116, 894-898.
- HAMAMATSU PHOTONICS. (2003). *PMA software U6039-01 INSTRUCTION MANUAL*
- Jiménez, B., & Carpio, A. (2008). *LA CATA DE ACEITES: ACEITE DE OLIVA VIRGEN. CARACTERÍSTICAS ORGANOLÉPTICAS Y ANÁLISIS SENSORIAL*. (Junta de Andalucía. Instituto de Investigación y Formación Agraria y Pesquera, Consejería de Agricultura y Pesca. ed.). Andalucía:
- Le Maire, G., Francois, C., & Dufrene, E. (2004). Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment*, 89(1), 1-28.
- Lleo, L., Roger, M., Herrero- Langreo, A., Diezma-Iglesias, B., & Barreiro, P. (2008). *Comparison of multispectral indexes extracted from hyperspectral images for the assessment of fruit ripening*. Madrid:
- Lozano Sanchez, J., Segura Carretero, A., & Fernandez Gutierrez, A. (2009). Composicion del aceite de oliva. In Unicaja (Ed.), *Aceite de oliva: Tesoro de andalucía* (pp. 194 final: 222)
- MAGRAMA. (2013). **Aceite de oliva y aceituna de mesa..** *Anuario De Estadística Agroalimentaria. Ministerio De Agricultura, Alimentación Y Medio Ambiente.*, (<http://www.magrama.gob.es/es/agricultura/temas/producciones-agricolas/aceite-oliva-y-aceituna-mesa/aceite.aspx>)
- Medina, J. (2013). *IMPLEMENTACIÓN DE ANÁLISIS MULTIVARIADO EN RESONANCIA MAGNÉTICA NUCLEAR PARA ESTUDIOS METABONÓMICOS* (UNIVERSIDAD DEL VALLE FACULTAD DE CIENCIAS NATURALES Y EXACTAS DEPARTAMENTO DE QUIMICA PROGRAMA DE QUÍMICA ed.). Santiago de Cali:
- Millar, S. J., Whitworth, M. B., & Chau, A. (2008). Mapping food composition using NIR hyperspectral imaging. *New Food Magazine*, Vol. 3
- Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46(2), 99-118.
- Oliveras López, M. J. (2005). Calidad del aceite de oliva virgen extra: Antioxidantes y función biológica.
- Otto, M. (2007). *Chemometrics statistics and computer application in analytical chemistry* (2a ed. ed.). Germany: WILEY-VCH Verlag GmbH & Co. KGaA.

- Picornell, M. R., & Melero, J. M. (2014). Historia del cultivo del olivo y el aceite; su expresión en la biblia [the history of oilk and the growig of olive tree; as expressed in the bible]. *ENSAYOS.Revista De La Facultad De Educación De Albacete*, 28, 155-181.
- Rojo, J. (2007). Regresión lineal múltiple. *IdEyGdM-Ld Estadística, Editor*, , 32.
- Salguero-Chaparro, L., Baeten, V., Fernández-Pierna, J. A., & Peña-Rodríguez, F. (2013). Near infrared spectroscopy (NIRS) for on-line determination of quality parameters in intact olives. *Food Chemistry*, 139(1), 1121-1126.
- Uceda, M., Hermoso, M., & Aguilera, M. P. (1998). La calidad del aceite de oliva. *El Cultivo Del Olivo*, , 547-572.

## VÍNCULOS

1. Producciones agrícolas - Aceite de oliva  
<http://www.magrama.gob.es/es/agricultura/temas/producciones-agricolas/aceite-oliva-y-aceituna-mesa/aceite.aspx>
2. Superficie y producción del cultivo del olivar MAGRAMA  
[http://www.magrama.gob.es/es/agricultura/temas/producciones-agricolas/2014\\_02\\_27\\_datos\\_WEB\\_olivar-final\\_tcm7-318770.pdf](http://www.magrama.gob.es/es/agricultura/temas/producciones-agricolas/2014_02_27_datos_WEB_olivar-final_tcm7-318770.pdf)
3. Regresión de mínimos cuadrados parciales  
<http://support.minitab.com/es-mx/minitab/17/topic-library/modeling-statistics/regression-and-correlation/partial-least-squares-regression/what-is-partial-least-squares-regression/>
4. Artículo espectroscopia del infrarrojo cercano.  
[http://www.engormix.com/MA-balanceados/formulacion/articulos/espectroscopia-infrarrojo-cercano-nirs-t577/800-p0.htm#\\_](http://www.engormix.com/MA-balanceados/formulacion/articulos/espectroscopia-infrarrojo-cercano-nirs-t577/800-p0.htm#_)
5. No solo de la Acidez vive el Aceite de Oliva – CSR Laboratorio.  
[http://www.csrservicios.es/joomla/index.php?option=com\\_content&view=article&id=110](http://www.csrservicios.es/joomla/index.php?option=com_content&view=article&id=110)
6. Cochinilla violeta en Olivo o parlatoria - Parlatoria olea  
<http://www.agroes.es/cultivos-agricultura/cultivos-frutales-y-fruticultura/olivo/cultivo-del-olivo/1236-cochinilla-violeta-en-olivo-o-parlatoria-parlatoria-oleae>

## 10 ANEXOS